# রাশিবিজ্ঞানের মূলতত্ত্ব

## (Basic Principles of Statistics)

প্রথম খণ্ড

( দুইখণ্ডে সম্পূর্ণ )


ডঃ শৈলেশভূষণ চৌধুরী এম্‌. এস্‌. সি., পি. এইচ্‌. ডি.

রাশিবিজ্ঞান বিভাগ, আশুতোষ কলেজ, কলকাতা।

ডঃ অরিজিৎ চৌধুরী এম্‌. এ., পি. এইচ্‌. ডি.

রাশিবিজ্ঞান বিভাগ, কলকাতা বিশ্ববিদ্যালয়।

শ্রীবিশ্বনাথ দাস এম্‌. এ.

রাশিবিজ্ঞান বিভাগ, প্রেসিডেন্সি কলেজ, কলকাতা।





পশ্চিমবঙ্গ রাজ্য পুস্তক পর্ষদ

( পশ্চিমবঙ্গ সরকারের একটি সংস্থা )

# উৎসর্গ

*স্বর্গত পিতৃদেব ও মাতৃদেবীর স্মৃতির উদ্দেশে*

শৈলেশভূষণ চৌধুরী

*মাতৃদেবীকে ও স্বর্গত পিতৃদেবের স্মৃতির উদ্দেশে*

অরিজিৎ চৌধুরী

*মাতৃদেবীকে ও স্বর্গত পিতৃদেবের স্মৃতির উদ্দেশে*

বিশ্বনাথ দাস

# মুখবন্ধ

বেশ কিছুদিন হ'ল আমাদের দেশে ইংরাজীর পাশাপাশি মাতৃভাষাকেও পাস পাঠক্রম স্নাতক স্তর পর্যন্ত শিক্ষাদানের মাধ্যম হিসাবে স্বীকার ক'রে নেওয়া হয়েছে। অতি সম্প্রতি মাতৃভাষার এই স্বীকৃতি সাম্মানিক স্নাতক ও স্নাতকোত্তর পর্যায়েও সম্প্রসারিত করা হয়েছে। কিন্তু দুঃখের বিষয়, রাশিবিজ্ঞানের বাংলা ভাষাভাষী ছাত্রছাত্রীগণ এই সুযোগ এখনও পাচ্ছে না, কারণ উল্লিখিত স্তরের ছাত্রছাত্রীদের উপযোগী বাংলা ভাষায় লিখিত রাশিবিজ্ঞানের পাঠ্যপুস্তক নেই। তাই পশ্চিমবঙ্গ রাজ্য পুস্তক পর্যদের কাছ থেকে এই গ্রন্থ প্রণয়নের দায়িত্ব পেয়ে উৎসাহিত বোধ করলেও এ কাজে উদ্যোগী হবার সমস্যা ভেবে আমাদের যথেষ্ট দ্বিধা ও সঙ্কোচ ছিল। কিন্তু এটা ঠিক যে অন্ততঃ প্রথম পর্যায়ে বিদেশী ভাষার মাধ্যম ছাত্রছাত্রীদের পক্ষে রাশিবিজ্ঞানের মত একটি অপেক্ষাকৃত নতুন বিষয় আয়ত্ত করার পথে একটি বড়সড় বাধা। রাশিবিজ্ঞানের শিক্ষক হিসাবে আমাদের এই অভিজ্ঞতা শেষ পর্যন্ত 'রাশিবিজ্ঞানের মূলতত্ত্ব' প্রণয়নের দুরূহ কাজে হাত দিতে আমাদের প্রেরণা জুগিয়েছে। তা ছাড়া প্রত্যেক নতুন উদ্যোগ এক সময় কাউকে না কাউকে তো শুরু করতেই হয়।

'রাশিবিজ্ঞানের মূলতত্ত্ব' প্রধানতঃ পশ্চিমবাংলার বিশ্ববিদ্যালয়গুলির স্নাতক পাঠক্রমের উপযোগী করে লেখা হয়েছে। তবে সাম্মানিক রাশিবিজ্ঞান, গণিতশাস্ত্র, উচ্চতর নিরীক্ষাশাস্ত্র, অর্থনীতি ও বাণিজ্যের ছাত্রছাত্রীদের অনেক প্রয়োজনও এই পুস্তকখানির সাহায্যে মিটতে পারে ব'লে আমাদের মনে হয়। পশ্চিমবাংলায় অধুনা প্রবর্তিত উচ্চতর মাধ্যমিক পাঠক্রমের ছাত্রছাত্রীদের পক্ষেও পুস্তকখানি বিশেষ উপযোগী হবে। এ ছাড়াও বিজ্ঞানের বিভিন্ন শাখায় যাঁরা গবেষণা করেন এবং পেশাগত প্রয়োজনে যাঁরা প্রতিনিয়ত রাশিবিজ্ঞানসম্মত বিভিন্ন পদ্ধতি প্রয়োগ করেন, তাঁদের পক্ষেও পুস্তকখানি অন্ততঃ আংশিকভাবে প্রয়োজনীয় বিবেচিত হতে পারে ব'লে আমাদের ধারণা। এই পুস্তক পাঠের পক্ষে বিদ্যালয়পাঠ্য গণিতের জ্ঞানই সাধারণভাবে পর্যাপ্ত হবে। তবে কয়েকটি পরিচ্ছেদে ম্যাট্রিক্স গণিত এবং অন্তরকলন ও সমাকলনের প্রাথমিক জ্ঞান প্রয়োজন হবে মনে রেখে পরিশিষ্টে এসম্বন্ধে কিছুটা আলোচনা করা হয়েছে।

পুস্তকখানি দুটি খণ্ডে প্রকাশিত হচ্ছে। প্রথম খণ্ডে (প্রথম থেকে একাদশ পরিচ্ছেদ পর্যন্ত) মোটামুটিভাবে রাশিতথ্য বিশ্লেষণের বিভিন্ন পদ্ধতি, প্রাথমিক সম্ভাবনাতত্ত্ব এবং দ্বিতীয় খণ্ডে (দ্বাদশ পরিচ্ছেদ থেকে শেষ পর্যন্ত) রাশিবিজ্ঞান-ভিত্তিক অনুমানতত্ত্ব সম্বন্ধে আলোচনা করা হয়েছে। পরিশিষ্টাংশটুকু থাকছে দ্বিতীয় খণ্ডে। প্রথম পরিচ্ছেদে রাশিবিজ্ঞানের সংজ্ঞা, প্রকৃতি, উদ্দেশ্য, উপযোগিতা ও সম্ভাব্য অপব্যবহার সম্পর্কে আলোচনা করা হয়েছে। পরবর্তী দুটি পরিচ্ছেদের বিষয়সূচীতে আছে রাশিতথ্য আহরণ, সারণী, লেখ ও চিত্রযোগে রাশিতথ্য উপস্থাপনার বিভিন্ন পদ্ধতি এবং পরিসংখ্যা বিভাজনের সাহায্যে রাশিতথ্য সংক্ষেপীকরণ সম্বন্ধে নানাবিধ আলোচনা। চতুর্থ থেকে ষষ্ঠ পরিচ্ছেদে অন্তর্ভুক্ত করা হয়েছে পরিসংখ্যা বিভাজনের মধ্যগামিতা, বিস্তৃতি, প্রতিবৈষম্য, তীক্ষ্ণতা, পরিঘাত, পরিসংখ্যারেখা প্রভৃতি বিষয়। এই পর্যায়ে সমগ্রক ও অংশকের মধ্যে পার্থক্য খুব একটা গুরুত্বপূর্ণ মনে না হওয়ায় এযাবৎ আলোচনা মোটামুটিভাবে নমুনালব্ধ রাশিতথ্যের ওপরই সীমাবদ্ধ রাখা হয়েছে। সপ্তম পরিচ্ছেদের বিষয়সূচী প্রাথমিক সম্ভাবনাতত্ত্ব। অষ্টম পরিচ্ছেদে বলা হয়েছে বিভিন্ন এক্‌চল তত্ত্বগত বিভাজন সম্বন্ধে। নবম থেকে একাদশ পরিচ্ছেদে গুণলক্ষণের সংস্রব, চলের সহগতি ও নির্ভরণ, মানক্রমিক সহগাঙ্ক, অন্তঃশ্রেণীক সহগাঙ্ক ইত্যাদি বিস্তারিতভাবে আলোচিত হয়েছে। দ্বাদশ পরিচ্ছেদে দেওয়া হয়েছে সাযুজ্যরেখা নিরূপণের বিভিন্ন পদ্ধতি। ত্রয়োদশ থেকে পঞ্চদশ পরিচ্ছেদে রাশিবিজ্ঞানভিত্তিক অনুমানতত্ত্ব স্থান পেয়েছে। এর মধ্যে ত্রয়োদশ পরিচ্ছেদে আছে নমুনাতত্ত্ব সম্বন্ধে প্রাথমিক আলোচনা এবং কিছু প্রয়োজনীয় নমুনাজ বিভাজন। প্রাক্‌কলন ও প্রকল্প-বিচারের মূলনীতি, নর্ম্যাল বিভাজন-ভিত্তিক কিছু যথার্থ প্রাক্‌কলন ও প্রকল্পবিচার এবং প্রভেদ-বিশ্লেষণ পদ্ধতি বর্ণিত হয়েছে চতুর্দশ পরিচ্ছেদে। আর পঞ্চদশ পরিচ্ছেদে আলোচিত হয়েছে বৃহৎ-নমুনাভিত্তিক আসন্নীকরণের উপযোগিতা এবং তার ওপর নির্ভরশীল কিছু প্রকাশন ও প্রকল্প-বিচার। পরিশিষ্টে ম্যাট্রিক্স গণিত, অন্তরকলন-সমাকলনের প্রাথমিক আলোচনা ছাড়াও আছে ভ্রান্তিতত্ত্ব, সংখ্যাভিত্তিক গণিত ইত্যাদি।

বিষয়বস্তু সহজবোধ্য করার জন্য সাধ্যমত উদাহরণ এবং চিত্রসহযোগে আলোচনার চেষ্টা করা হয়েছে। যথাসম্ভব বাস্তব ও ভারতীয় রাশিতথ্য ব্যবহারের সাহায্যে পুস্তকটিকে আকর্ষণীয় ক'রে তোলার দিকে লক্ষ্য রাখা হয়েছে। ছাত্রছাত্রীদের অধীতবিদ্যা চর্চার সুবিধার জন্য প্রতি পরিচ্ছেদের শেষে

বেশ কিছু সুনির্বাচিত প্রশ্ন দেওয়া হয়েছে। এ ছাড়া আগ্রহী পাঠক-পাঠিকাদের জন্য বিভিন্ন বিষয়ের ওপর নির্বাচিত পুস্তক-তালিকাও দেওয়া হয়েছে।

বাংলাভাষায় এই পুস্তক প্রণয়ণের কাজ হাতে নিয়ে আমাদের সবচেয়ে বেশী যে অসুবিধার সম্মুখীন হতে হয়েছে সেটা হ'ল উপযুক্ত পরিভাষার অভাব। এ ব্যাপারে আমরা মোটামুটিভাবে ডঃ পূর্ণেন্দুকুমার বসুর 'রাশিবিজ্ঞানের গোড়ার কথা' ( বিশ্বভারতী, 1956 ) এবং শ্রীভাগবত দাশগুপ্ত, ডঃ অরিজিৎ চৌধুরী ও শ্রীবিশ্বনাথ দাস সঙ্কলিত 'রাশিবিজ্ঞানের পরিভাষা' ( পশ্চিমবঙ্গ রাজ্য পুস্তক পর্ষদ, 1972 ) পুস্তিকা-দুটির ওপর নির্ভর করেছি। অত্যন্ত বিশেষ অর্থে ব্যবহৃত এবং আন্তর্জাতিক স্বীকৃতিসম্পন্ন শব্দ ভাষান্তরিত করা হয়নি এবং বিজ্ঞানের অন্যান্য শাখায় গৃহীত পরিভাষা যথাসম্ভব অবিকৃত রাখা হয়েছে। প্রাথমিক অসুবিধার কথা স্মরণ রেখে কোন পরিভাষা এই পুস্তকে প্রথমবার ব্যবহারের সময় বন্ধনীতে ইংরাজী প্রতিশব্দটি দেওয়া হয়েছে। ব্যবহৃত পরিভাষা প্রামাণ্য ব'লে আমরা দাবী করি না—কিছু কিছু পরিভাষার উন্নতিসাধনের অবকাশ নিশ্চয়ই আছে। শিক্ষক, গবেষক, ছাত্র ও সাধারণ পাঠকবৃন্দের কাছ থেকে এই পুস্তক সম্পর্কিত সুচিন্তিত মতামত ও পরামর্শ আহ্বান করছি। ভবিষ্যৎ মুদ্রণে প্রয়োজনবোধে তদনুযায়ী পুস্তকটির পরিবর্তনসাধনে আমরা সচেষ্ট হব।

এই পুস্তকখানি প্রণয়নে উৎসাহ ও পরামর্শ দিয়ে এবং আরও নানাভাবে আমাদের সাহায্য করেছেন ডঃ পূর্ণেন্দুকুমার বসু, শ্রীঅনিলকুমার ভট্টাচার্য, শ্রীহরিকিঙ্কর নন্দী, স্বর্গত ডঃ অনুকুলচন্দ্র দাস প্রমুখ বিশিষ্ট রাশিবিজ্ঞানীগণ, পশ্চিমবঙ্গ রাজ্য পুস্তক পর্ষদের রাশিবিজ্ঞান বিষয়ক সমিতির অন্যান্য সদস্যবৃন্দ এবং আমাদের বিভিন্ন সহকর্মী ও বন্ধুগণ। এই প্রসঙ্গে শ্রীদীপংকর বসুর নাম বিশেষভাবে উল্লেখযোগ্য। ডঃ অতীন্দ্রমোহন গুণ প্রথম পর্যায়ে সমগ্র পাণ্ডুলিপিখানি আদ্যোপান্ত পুঙ্খানুপুঙ্খরূপে পাঠ ক'রে যে সব মূল্যবান মতামত দিয়েছেন সেগুলি পুস্তকখানির উৎকর্ষবিধানে যথেষ্ট সাহায্য করেছে। দৈনিক স্টেট্‌স্‌ম্যান পত্রিকা, ইণ্ডিয়ান ফুটবল অ্যাসোসিয়েশন এবং হুগলী জেলার ইছাপুর উচ্চ বিদ্যালয় ও ইছাপুর পাবলিক লাইব্রেরীর কর্তৃপক্ষ কিছু প্রয়োজনীয় রাশিতথ্য সরবরাহ ক'রে আমাদের সহায়তা করেছেন। এঁদের সকলকে আমাদের আন্তরিক কৃতজ্ঞতা জানাই। আর ধন্যবাদ জানাই পশ্চিমবঙ্গ রাজ্য পুস্তক পর্ষদের সদস্যদের, বিশেষ ক'রে মুখ্য প্রশাসন আধিকারিক শ্রীঅবনী মিত্রকে, যাঁদের

উদ্যোগে এই পুস্তকখানি প্রকাশ করা সম্ভবপর হয়েছে এবং কে. পি. বসু প্রিন্টিং ওয়ার্কস-এর কর্তৃপক্ষ ও কর্মিবৃন্দকে, যাঁদের যত্ন, শ্রম ও কৃতিত্বে পুস্তকটির মুদ্রণ-সৌকর্য আশানুরূপ স্তরে পৌঁছেছে।

গ্রন্থখানি পাঠকসমাজে সমাদৃত হলে আমাদের শ্রম সার্থক বিবেচিত হবে।

কলকাতা
জুলাই, 1976

শৈলেশভূষণ চৌধুরী
অরিজিৎ চৌধুরী
বিশ্বনাথ দাস

# সূচীপত্র

## প্রথম খণ্ড

# অবতরণিকা
# ( Introduction ) 1

## 1.1 রাশিবিজ্ঞান এবং পরিসংখ্যান :

ইংরেজী Statistics কথাটির উৎপত্তি State অর্থাৎ রাষ্ট্র থেকে। রাষ্ট্র-শাসনের প্রয়োজনে প্রাচীনকালে যে-সমস্ত তথ্য আহরণ করা হ'ত সেগুলিকে সাধারণভাবে বলা হয় Statistics—যেমন, জনসংখ্যা, সামরিক শক্তি-সংক্রান্ত তথ্য, আদায়ীকৃত রাজস্বের পরিমাণ, ইত্যাদি। পরবর্তীকালে অবশ্য কথাটি আরও ব্যাপকতর অর্থে ব্যবহার করা হচ্ছে। যে-দুটি বিভিন্ন অর্থে বর্তমান Statistics কথাটির প্রচলন সে-দুটির সঙ্গে সঙ্গতি রেখে আমরা এর দুটি প্রতিশব্দ ব্যবহার করব—**পরিসংখ্যান** এবং **রাশিবিজ্ঞান।**

শুধুমাত্র রাষ্ট্রশাসনের সূত্রে সংগৃহীত রাশিতথ্যই নয়, যে-কোন বিশেষ উদ্দেশ্যে পার্থিব যে-কোন ঘটনা সম্পর্কে সংগৃহীত রাশিতথ্যকেই বর্তমানে Statistics বলা হয়ে থাকে। এই অর্থে আমরা 'পরিসংখ্যান' প্রতিশব্দটি ব্যবহার করব। সাধারণ মানুষের কাছে Statistics কথাটি এই অর্থেই বেশী পরিচিত—যেমন আমরা ব'লে থাকি, জনস্বাস্থ্য-সংক্রান্ত পরিসংখ্যান, বিগত তিন দশকে দেশে খাদ্যশস্য উৎপাদনের পরিসংখ্যান, ইত্যাদি।

পরিসংখ্যান বা রাশিতথ্য সংগ্রহ করার পশ্চাতে সব সময়েই একটি উদ্দেশ্য থাকে—এই উদ্দেশ্য হচ্ছে, যে বিশেষ ঘটনাটির ওপর রাশিতথ্য সংগৃহীত হ'ল সেটি বিশেষ কোন্ কোন্ কারণের ফলশ্রুতি, অথবা সংগৃহীত রাশিতথ্য সাধারণভাবে কোন্ সত্যটির ইঙ্গিতবহ, তা খুঁজে বের করা। সাধারণতঃ এই সব কার্যকারণ সম্পর্কগুলি জটিল এবং নিয়ন্ত্রণবহির্ভূত হয়ে থাকে, তাই রাশিতথ্য আহরণের পর তা উপযুক্তভাবে উপস্থাপন, বিশ্লেষণ এবং ব্যাখ্যানের প্রশ্ন দেখা দেয় অনিবার্যভাবে। যে শাস্ত্র রাশিতথ্য আহরণ, উপস্থাপন, বিশ্লেষণ এবং ব্যাখ্যানের বিজ্ঞান-সম্মত পদ্ধতি নিয়ে আলোচনা করে সেটিকে আমরা অভিহিত করব 'রাশিবিজ্ঞান' নামে। লক্ষণীয়, ইংরেজী Statistics কথাটি এই অর্থেও প্রচলিত।

এই দুটি ভিন্ন অর্থে ইংরেজীতে কেবলমাত্র Statistics কথাটিরই ব্যবহার

হয়। এর বাংলা প্রতিশব্দ-দুটির অর্থের পার্থক্য সব সময় মনে রাখতে হবে। ইংরেজী Statistician কথাটির প্রতিশব্দ রাশিবিজ্ঞানী—অর্থাৎ যিনি রাশিবিজ্ঞানশাস্ত্রে পারদর্শী, এমন নয় যে, বিভিন্ন ধরনের পরিসংখ্যান তার নখদর্পণে।

## 1.2. রাশিবিজ্ঞানের প্রকৃতি এবং উদ্দেশ্য :

তাহলে দেখা যাচ্ছে, রাশিবিজ্ঞানকে একটি শাস্ত্র আখ্যা দেওয়া হল। এখন প্রশ্ন হতে পারে, রাশিবিজ্ঞান কি বিজ্ঞানের একটি শাখা, নাকি প্রকৃতিতে এটি একটি কলাবিশেষ ? নামের সঙ্গে সঙ্গতি রেখে অবশ্য একে বিজ্ঞান বলাই যুক্তিযুক্ত হবে, তবে বিজ্ঞানের প্রচলিত শাখাগুলির সঙ্গে এর প্রকৃতিগত একটি মৌলিক পার্থক্য রয়েছে। প্রচলিত প্রাকৃতিক বিজ্ঞানগুলি কিছু কিছু বিধির সমষ্টিবিশেষ। এই বিধিগুলি প্রথমতঃ বিজ্ঞানের সংশ্লিষ্ট শাখার বৈজ্ঞানিকদের মনে অনুমানের (conjecture) আকারে জন্ম নেয়। অনুমানের উপর ভিত্তি ক'রে একটি প্রকল্প (hypothesis) রচনা ক'রে অতঃপর শুরু হয় অনুমানটি নিয়ে পরীক্ষা-নিরীক্ষা। পরীক্ষালব্ধ ফল অনুমানের সপক্ষে গেলে অনুমানটি উন্নীত হয় বিধিতে, অন্যথায় এটি যায় বাতিল হয়ে। এখন বিজ্ঞান হিসাবে রাশিবিজ্ঞানের প্রকৃতি ঠিক এই ধরনের নয়। বরঞ্চ বলা চলে রাশিবিজ্ঞান বিজ্ঞানের অন্যান্য শাখাকে অনুমানলব্ধ প্রকল্প থেকে বিধিতে উত্তরণে বিজ্ঞানসম্মত পদ্ধতির হদিশ দেয়। এখন, এই সব পদ্ধতিগুলি বিজ্ঞানসম্মত, সুতরাং সেই অর্থে রাশিবিজ্ঞানকে বিজ্ঞান আখ্যা দেওয়া চলতে পারে।

একটি উদাহরণ নেওয়া যাক। মনে কর, জনৈক কৃষিবিজ্ঞানী অনুমান করলেন, একটি বিশেষ ধরনের সার বাজারে প্রচলিত অন্যান্য সারের তুলনায় ধানচাষের পক্ষে অনেক বেশী উপযোগী হবে। তিনি পরীক্ষা-নিরীক্ষা চালালেন। অনুরূপ পরিস্থিতিতে বিশেষ এই সারটির প্রয়োগে উৎপাদনের পরিমাণ প্রচলিত অন্যান্য সার প্রয়োগে উৎপাদনের পরিমাণের সঙ্গে তুলনা করা হ'ল। একাধিক পরীক্ষার ফলাফলে প্রকৃতিগত এবং পরিমাণগত পার্থক্য থাকা খুবই স্বাভাবিক। এক্ষেত্রে এই সমস্ত ফলাফল একত্রিত ক'রে কি-ভাবে একটি সিদ্ধান্তে আসা যায়, রাশিবিজ্ঞানসম্মত নানান পদ্ধতি প্রয়োগ ক'রে তার সন্ধান মেলে। অর্থাৎ, পরীক্ষালব্ধ সীমিতসংখ্যক তথ্যের অন্তর্নিহিত বৈষম্য বিশ্লেষণ ক'রে কতখানি সাধারণ সত্য আহরণ করা যায় এবং সঙ্গত কী সিদ্ধান্তে উপনীত হওয়া যায়, তাই হ'ল রাশিবিজ্ঞানের প্রধান উপজীব্য বিষয়।

এখন প্রশ্ন হ'ল, কোন্ মূল নীতির নিরিখে রাশিবিজ্ঞানসম্মত পদ্ধতিতে এইসব প্রশ্নের বিচার হবে ? 'ক' পরিস্থিতির অবতারণায় যদি প্রতিবারই 'খ' ফলটির উদ্ভব হয় তাহলে 'ক' যে 'খ'-এর কারণ—সে সিদ্ধান্তে পৌঁছানোর জন্য রাশিবিজ্ঞানের অপেক্ষা করবেন না কেউই। সুতরাং যে রাশিতথ্যে বৈচিত্র্যের অভাব তা রাশিবিজ্ঞানের আওতায় আসে না। পক্ষান্তরে যে-সব রাশিতথ্যে বৈচিত্র্যের আভাস, সেখানেই প্রয়োজন হয় রাশিবিজ্ঞানের। 100টির মধ্যে 97টি ক্ষেত্রে 'ক'-এর ফলশ্রুতি 'খ' হলে বাকী তিনটিতে না হলেও 'ক'-কে 'খ'-এর কারণ বলা চলবে কি না, অথবা কিছু সংখ্যক ভারতীয়ের গড় উচ্চতা 64 ইঞ্চি লক্ষ্য ক'রে সাধারণভাবে ভারতীয়দের গড় উচ্চতা 62 থেকে 66 ইঞ্চির মধ্যে হবে—একথা বলা যাবে কিনা, বা গেলে কতখানি আস্থার সঙ্গে বলা যাবে, কিংবা পরস্পর সম্বন্ধযুক্ত দুটি লক্ষণের ( যেমন মনে কর, উচ্চতা এবং ওজন ) একটির মান বিশেষ ক্ষেত্রে জানা থাকলে অন্যটির সম্বন্ধে কতখানি নিশ্চয়তার সঙ্গে বলা যাবে—এইসব প্রশ্নের রাশিবিজ্ঞানসম্মত বিচার হয় **সম্ভাবনাতত্ত্বের** (Theory of Probability) ভিত্তিতে। আসলে সাধারণভাবে জাগতিক সমস্ত ঘটনাই একটি বিশেষ নিয়মের (Law of Uniformity) অধীন হলেও পরিস্থিতি এবং পরিবেশের বৈচিত্র্যের দরুণ ফলশ্রুতিতেও বৈচিত্র্য অনিবার্য। তাই সাম্প্রতিকতম মতবাদ অনুযায়ী বিজ্ঞানসম্মত কোন বিধির সঠিক বয়ান 'নির্দিষ্টভাবে ক খ-এর কারণ' না হয়ে হওয়া উচিত 'ক খ-এর কারণ হওয়ার সম্ভাবনা খুব বেশী'। রাশিবিজ্ঞান কেবলমাত্র এই আকারেই বিধি প্রতিষ্ঠা করতে বিজ্ঞানের অন্যান্য শাখাকে সাহায্য করে।

সাধারণভাবে রাশিবিজ্ঞানের উপজীব্য 'তথ্যসমষ্টি'—একক পরিস্থিতিতে ব্যক্তিবিশেষ বা বস্তুবিশেষ সংক্রান্ত তথ্যে রাশিবিজ্ঞান পৃথক্‌ভাবে আগ্রহী নয়। যেমন, বিশ্ববিদ্যালয়ের একটি পরীক্ষায় গড়ে কতজন উত্তীর্ণ হয়েছে এই ধরনের তথ্যানুসন্ধানের সূত্রেই কেবল জনৈক শ্রীমান ক-এর উক্ত পরীক্ষায় প্রাপ্ত নম্বরটিতে রাশিবিজ্ঞানী আগ্রহী হবেন—পৃথক্‌ভাবে এই নম্বরের কোন গুরুত্বই তার কাছে নাই। তেমনি দিনের একটি বিশেষ সময়বিন্দুতে শ্রীমতী খ-এর দৈহিক তাপমাত্রা রাশিবিজ্ঞানের আওতায় তখনই আসবে যখন রোগাক্রান্ত শ্রীমতী খ-এর দৈহিক তাপমাত্রার গতিধারা (trend) বিশ্লেষণ-জাতীয় প্রশ্ন দেখা দেবে, অন্যথায় নয়।

সমষ্টির কোন লক্ষণের উপর সংগৃহীত রাশিতথ্য নিয়েই সাধারণতঃ রাশি-

বিজ্ঞানের বিচার-বিশ্লেষণ চললেও অধিকাংশ ক্ষেত্রেই মূল লক্ষ্য থাকে বৃহত্তর কোন সমষ্টির (প্রথম সমষ্টিটি যার একটি অংশবিশেষ) সংশ্লিষ্ট লক্ষণটির উপর আলোকপাত করা। যেমন মনে কর, আমরা কলকাতার অধিবাসীদের মাসিক আয়ের গড় পরিমাণ নির্ণয় করতে চাই। যথার্থ উত্তরটি পেতে হলে কলকাতার প্রতিটি অধিবাসীর কাছে উপস্থিত হয়ে তথ্যসংগ্রহ করায় যে পরিমাণ শ্রম, অর্থ এবং সময় প্রয়োজন তা সঙ্কুলান করা অনেক সময় আমাদের সাধ্যাতীত হয়ে পড়ে। বিকল্পভাবে সীমিত-সংখ্যক ( ধরা যাক 500, কিংবা 1,000 ) অধিবাসী-দের কাছ থেকে পাওয়া তথ্যের উপর ভিত্তি ক'রে বৈজ্ঞানিক পদ্ধতিতে আমাদের আলোচ্য প্রশ্নটির যথাযোগ্য সমাধানে রাশিবিজ্ঞান আমাদের সাহায্য করতে পারে। কোন **সমগ্রকের** (population) এই ধরনের অংশবিশেষকে **নমুনা** (sample) বলা হয়। স্পষ্টতঃই যে নমুনাতে সমগ্রকের মূল বৈশিষ্ট্যগুলি যতখানি বিশ্বস্ততার সঙ্গে রক্ষিত হবে, অর্থাৎ যে নমুনা যত বেশী প্রতিনিধি-স্থানীয় হবে, সেটি এই প্রসঙ্গে তত কার্যকরী হবে। ব্যষ্টি থেকে সমষ্টিতে, নমুনা থেকে সমগ্রকে উত্তরণের এই পদ্ধতিটি হ'ল **আরোহী অনুমান** (Inductive Inference) **পদ্ধতি**। রাশিবিজ্ঞানের ভিত্তি হ'ল মূলতঃ এই পদ্ধতিটি।

বিজ্ঞানের বিভিন্ন শাখায় রাশিবিজ্ঞানের প্রয়োগ ক্রমশঃ ব্যাপকতর হচ্ছে। বিশেষ বিশেষ ক্ষেত্রে রাশিবিজ্ঞানসম্মত পদ্ধতিগুলির প্রয়োগে জন্ম নিয়েছে **ফলিত রাশিবিজ্ঞানের** (Applied Statistics) বিবিধ শাখা, সুতরাং এক অর্থে রাশিবিজ্ঞান শাস্ত্রটি একটি কলাবিশেষও বটে।

বর্তমান গ্রন্থে রাশিবিজ্ঞানের কতকগুলি বিজ্ঞানসম্মত পদ্ধতি নিয়ে আলোচনা করা হবে। যথার্থ অর্থে এটি বিজ্ঞান, অথবা কলা অথবা উভয়ই—সেই বিস্তারিত বিতর্কে আমরা অধিক অগ্রসর হব না।

## 1.3 রাশিবিজ্ঞানের উপযোগিতা :

রাশিবিজ্ঞানে পর্যাপ্ত জ্ঞানের অভাবে সমাজ-বিজ্ঞানের যে-কোন শাখার একজন গবেষকের অবস্থাখানি 'নিশ্ছিদ্র অন্ধকারময় একটি কক্ষে অনুপস্থিত একটি কালো বিড়াল অন্বেষণরত একজন অন্ধের'* মত করুণ হয়ে পড়তে পারে। অর্থনীতি, সমাজবিদ্যা, শারীরবিদ্যা, প্রাণিবিদ্যা, ভেষজবিজ্ঞান, কৃষিবিজ্ঞান, মনোবিজ্ঞান, শিক্ষাবিজ্ঞান প্রভৃতি সমাজবিজ্ঞান ও জীববিজ্ঞানের বিভিন্ন শাখায়

* [1], পৃঃ 1.

রাশিবিজ্ঞানের বহুল ব্যবহার অনেককাল আগে থেকেই প্রচলিত। অধুনা, পদার্থবিত্তা, রসায়নবিত্তা প্রভৃতি তথাকথিত 'যথার্থ' বিজ্ঞানগুলির ক্ষেত্রেও রাশিবিজ্ঞানের উপযোগিতা ক্রমশঃ বেশী পরিমাণে স্বীকৃত হচ্ছে।

সরকারের কাছেও রাশিবিজ্ঞানের গুরুত্ব সামান্ত নয়। সুষ্ঠু প্রশাসন ব্যবস্থা প্রণয়নে, প্রয়োজনানুগ বাস্তবমুখী পরিকল্পনা রচনায় এবং বিবিধ নীতি-নির্ধারণে জনসংখ্যা, জনস্বাস্থ্য, প্রাকৃতিক সম্পদের পরিমাণ, আবহাওয়া, কৃষিজাত ও শিল্পজাত দ্রব্যের উৎপাদনের পরিমাণ, বেকারদের সংখ্যা, ইত্যাদি, ইত্যাদি হাজারো পরিসংখ্যান সংগ্রহ ও বিশ্লেষণের প্রয়োজনীয়তা উল্লেখের অবকাশ রাখে না।

শিল্পক্ষেত্রেও রাশিবিজ্ঞানের উপযোগিতা অপরিসীম। ক্রেতার চাহিদা অনুযায়ী উপযুক্ত মানের শিল্পসামগ্রী উৎপাদনে রাশিবিজ্ঞানসম্মত পদ্ধতি প্রয়োগ আজ অপরিহার্য হয়ে উঠেছে।

আজকের যুগটি দ্রুত শিল্পায়নের যুগ, তাই তীব্র প্রতিযোগিতারও যুগ। তাই **চাহিদা-সংক্রান্ত গবেষণা** (Market Research) আজ প্রথম শ্রেণীর গবেষণার বিষয়বস্তুর পর্যায়ে উন্নীত হয়েছে। এক্ষেত্রেও রাশিবিজ্ঞানের রয়েছে সফল ভূমিকা।

রাশিবিজ্ঞানের ব্যবহার বিভিন্ন ক্ষেত্রে এত দ্রুত প্রসার লাভ করছে যে বর্তমান স্বল্প পরিসরে সবগুলির উল্লেখ সম্ভবপর নয়। তবে রাশিবিজ্ঞানের সব থেকে বড় উপযোগিতা বোধ হয় আজকের দিনে একজন দায়িত্বশীল সচেতন নাগরিকের কাছে। বর্তমান যুগটি এক কথায় প্রচারের যুগ—নির্বাচন-প্রার্থী থেকে শুরু ক'রে দেশের সরকার পর্যন্ত নিজেদের অনুকূলে অবিরাম প্রচার চালাচ্ছেন অনর্গল রাশিতথ্যের উদ্ধৃতি দিয়ে। এখন রাশিতথ্য উপস্থাপনার, তথা রাশি-বিজ্ঞানের মূল নীতিগুলির সঙ্গে পরিচয় থাকলে একজন সাধারণ মানুষের পক্ষে রাশিতথ্যগুলি সঠিক অর্থে এবং পরিপ্রেক্ষিতে গ্রহণ করা সম্ভব হবে। সুতরাং সেক্ষেত্রে রাশিতথ্যের সাহায্যে প্রতারিত করার তথাকথিত 'সহজ' পথে তাকে প্রতারিত করা সহজে নাও সম্ভব হতে পারে।

## 1.4 পরিসংখ্যানের অপব্যবহার :

বিজ্ঞানের বিভিন্ন শাখার সহায়ক হিসাবে এবং অন্তান্ত নানান ক্ষেত্রে রাশি-বিজ্ঞানের জনপ্রিয়তা একদিকে যদিও ক্রমবর্ধমান, অন্তদিকে আবার একশ্রেণীর

সাধারণ মানুষের মনে পরিসংখ্যান তথা রাশিবিজ্ঞানের ওপর আস্থার একান্ত অভাব। Disraeli-র কালজয়ী উক্তিটি—"There are three kinds of lies—lies, damned lies and statistics" ( মিথ্যা তিনপ্রকার—মিথ্যা, নির্জলা মিথ্যা এবং পরিসংখ্যান )—এই প্রসঙ্গে সকলেরই মনে পড়বে। সাধারণ মানুষের পরিসংখ্যানের তথা রাশিবিজ্ঞানের ওপর এই ধরনের আস্থাহীনতার একটা বড় কারণ হ'ল, আমাদের দৈনন্দিন জীবনে প্রতিনিয়তঃ যেসব পরিসংখ্যান সংগৃহীত এবং উদ্ধৃত হয়, সংগ্রাহকের অভিজ্ঞতা, দক্ষতা এবং সর্বোপরি অনেক সময় সততার অভাবের দরুণ সেগুলিতে এত ভুল থাকে যে অধিকাংশ সময়েই এগুলি বাস্তবচিত্রের পরিবর্তে একটি ভ্রান্ত চিত্র কিংবা উদ্দেশ্যপ্রণোদিত চিত্র দিয়ে থাকে। আর একটা কারণ, অনেকের ধারণা পরিসংখ্যানের সাহায্যে যা খুশী তাই প্রমাণ করা যায়। এটি একটি গুরুত্বপূর্ণ অভিযোগ। অসাধু ব্যক্তিরা অনেক সময় নিজেদের উদ্দেশ্যসিদ্ধির জন্য পরিসংখ্যানের অপব্যবহার ক'রে থাকে ভুল পরিসংখ্যান উদ্ধৃত ক'রে, কিংবা পরিসংখ্যান ভুল পরিপ্রেক্ষিতে উপস্থাপন করে। অনেক সময় অজ্ঞতাবশতঃ এই ভুল ঘটে যায়। আসল কথা,—'Figures seldom lie, only liars figure'. পরিসংখ্যান সত্য প্রতিষ্ঠায় খুবই উপযোগী এবং গুরুত্বপূর্ণ, তবে এটির সঠিক ব্যাখ্যান প্রয়োজন এবং এটিকে গ্রহণ করতে হবে সঠিক পরিপ্রেক্ষিতে। কারণ পরিসংখ্যান নিজে থেকে কিছুই সূচিত করে না, পরিসংখ্যানকে সঠিক অর্থে গ্রহণ করলে তবেই তা থেকে প্রয়োজনীয় সত্য উদ্‌ঘাটিত হয়। নয়তো অসতর্ক, অদক্ষ এবং উদ্দেশ্যবিহীন কিংবা উদ্দেশ্যপ্রণোদিত ব্যবহারের ফলে পরিসংখ্যান মিথ্যাকেও সত্য হিসাবে উপস্থাপন করতে পারে। ফলে সাধারণ মানুষের আস্থা নষ্ট হয়ে যায় পরিসংখ্যান তথা রাশিবিজ্ঞানের ওপর।

পরিসংখ্যানের অপব্যবহারের কয়েকটি উদাহরণ নীচে আলোচনা করা হ'ল।

### অসম্পূর্ণ রাশিতথ্য থেকে সিদ্ধান্ত গ্রহণ :

ধরা যাক, হিসেব করে দেখা গেল, মদ্যপায়ীদের গড় আয়ু 45 বছর। সুতরাং সিদ্ধান্ত নেওয়া হ'ল মদ্যপান মানুষের আয়ুর পক্ষে ক্ষতিকারক। বাস্তবিকপক্ষে এই ধরনের সিদ্ধান্ত নেওয়ার আগে যারা মদ্যপান করে না তাদের গড় আয়ু সম্বন্ধে খোঁজ নেওয়া এবং দুটির মধ্যে তুলনা করার প্রয়োজন ছিল। তা করা হয় নি, সুতরাং সিদ্ধান্তটি বৈধ কিনা তা আজও বিচার সাপেক্ষ।

**অসম তুলন ঃ**

গতবছর দেশে সৈন্যবাহিনীতে মৃত্যুর সংখ্যা এবং দুর্ঘটনা-জনিত মৃত্যুর সংখ্যা দাঁড়াল ধরা যাক, যথাক্রমে 98 হাজার ও $95\frac{1}{2}$ হাজার—সুতরাং বলা হ'ল, বাড়িতে থাকার থেকে যুদ্ধে যাওয়া এমন কিছু বেশী বিপজ্জনক নয়। এখানে স্পষ্টতঃই উভয় কারণ থেকে 'মৃত্যুহার'-দুটি তুলনা করাই যুক্তিযুক্ত—'মৃত্যুসংখ্যা' নয়, কারণ সৈন্যবাহিনীর মোট লোকসংখ্যা থেকে অসামরিক জনসংখ্যা অনেক গুণে বেশী।

**শতকরা হার বা অনুপাতের ভুল ব্যবহার ঃ**

অনেক সময় শুধুমাত্র অনুপাত বা শতকরা হারের উল্লেখে ভ্রান্ত ধারণার সৃষ্টি হতে পারে। একটি স্কুলের শিক্ষক-শিক্ষিকাদের মধ্যে দুজন শিক্ষিকা। এঁদের মধ্যে হয়তো একজন ছাত্র-ছাত্রীদের কাছে জনপ্রিয় নন। দুজন শিক্ষিকার মধ্যে একজন জনপ্রিয় নন—এ তথ্যটি তেমন অস্বাভাবিক নয়, কিন্তু মোট শিক্ষিকার সংখ্যা উল্লেখ না ক'রে যদি কেবল বলা হয়, ঐ স্কুলের শতকরা পঞ্চাশজন শিক্ষিকা জনপ্রিয় নন, তাহলে তথ্যটি ভুল নয় ঠিকই। কিন্তু প্রথম দৃষ্টিতে এটি নিশ্চয়ই উদ্বেগের কারণ হবে।

**পরিবর্তনশীল গোষ্ঠী-সংক্রান্ত পরিসংখ্যান ব্যবহার ঃ**

কোন কলেজের প্রাক্তনী সংসদের সদস্যদের গড় বয়স 1969 সালে ছিল 56, কিন্তু 1970 সালে দাঁড়াল 54. তথ্যটি পড়ে প্রথম দৃষ্টিতে মনে হবে সদস্যদের বয়স বুঝি সত্যিই কমে যাচ্ছে। আসলে ব্যাপারটি হ'ল, উল্লিখিত দুবছরে প্রাক্তনী সংসদের গঠন এক নয়—1970 সালে বয়স্ক কিছু প্রাক্তনী মারা গিয়েছেন এবং অপেক্ষাকৃত অল্পবয়সী কিছু নতুন সদস্যের অন্তর্ভুক্তি ঘটেছে।

**ত্রুটিপূর্ণ সংজ্ঞা ব্যবহার ঃ**

1961 সালের আদমশুমারিতে কলকাতা এবং বোম্বাই এই দুটি শহরের জনসংখ্যা দেখানো হ'ল যথাক্রমে 2,927,289 এবং 4,152,056. দেখে মনে হবে, কলকাতার থেকে বোম্বাইয়ের জনসংখ্যা সত্যিই বুঝি বেশী। কিন্তু আসল তথ্য হ'ল 2,927,289 শুধুমাত্র কলকাতা কর্পোরেশন এলাকার জনসংখ্যা, কিন্তু 4,152,056 হচ্ছে বৃহত্তর বোম্বাই-এর জনসংখ্যা। সুতরাং এখানে শহরের সংজ্ঞা দুটি ক্ষেত্রে এক নয়।

**প্রতিনিধিস্থানীয় নয়, এমন নমুনা ব্যবহার :**

1,000টি নমুনা সমীক্ষা ক'রে জনৈক সমীক্ষক ঘোষণা করলেন, কলকাতা-বাসীদের শতকরা 55 জনের নিজেদের গাড়ি আছে। তথ্যটি নিঃসন্দেহে চাঞ্চল্যকর। কিন্তু পরে খোঁজ নিয়ে জানা গেল, সমীক্ষক ভদ্রলোক নমুনা সংগ্রহের ব্যাপারে টেলিফোন ডিরেক্টরির আশ্রয় নিয়েছিলেন। আসলে সাধারণতঃ কেবল সঙ্গতিসম্পন্ন ব্যক্তিদেরই টেলিফোন থাকে। সুতরাং গৃহীত নমুনাটি এখানে আদৌ প্রতিনিধিমূলক হয়নি।

**অপর্যাপ্ত রাশিতথ্য থেকে সিদ্ধান্ত গ্রহণ :**

আট-দশজন পরিচিত ধূমপায়ীকে ক্যান্সার (cancer) রোগে আক্রান্ত হতে দেখে ধূমপানকে ক্যান্সার রোগের কারণ হিসাবে বর্ণনা করা নিশ্চয়ই যুক্তিযুক্ত হবে না। এই ধরনের সিদ্ধান্তে আসতে হলে আরও অনেক বেশী সংখ্যক ধূমপায়ী এবং ক্যান্সার রোগী পর্যবেক্ষণ করতে হবে এবং গৃহীত নমুনা যাতে প্রতিনিধি-স্থানীয় হয় সেদিকেও লক্ষ্য রাখতে হবে।

পরিসংখ্যানের অপব্যবহারের এই ধরনের আরও অনেক উদাহরণ দেওয়া যেতে পারে। পরিসংখ্যান সঠিকভাবে ব্যবহার করতে শেখা যেমন গুরুত্বপূর্ণ, তেমনি প্রয়োজন পরিসংখ্যানের সম্ভাব্য অপব্যবহারের বিরুদ্ধে সতর্কতা অবলম্বন করা। শুধু রাশিবিজ্ঞানের ছাত্র বা কর্মীই নয়, আজকের যুগে বিজ্ঞান-সাধক, দক্ষ প্রশাসক, সফল রাজনৈতিক নেতা কিংবা সচেতন নাগরিক—কারোরই পরিসংখ্যানের সাহায্যে অযথা বিভ্রান্ত হওয়া চলে না। সুতরাং সকলকেই এ ব্যাপারে সচেতন হতে হবে।

## 1.5 অনুশীলনী

1.1 রাশিবিজ্ঞান ও পরিসংখ্যানের সংজ্ঞা দাও। রাশিবিজ্ঞানের উদ্দেশ্য ও প্রকৃতি বর্ণনা কর।

1.2 রাশিবিজ্ঞানের উপযোগিতা বর্ণনা কর। বিজ্ঞানের অন্যান্য শাখার সঙ্গে এই শাস্ত্রটির সম্পর্ক নির্দেশ কর।

1.3 রাশিবিজ্ঞানের ওপর সাধারণ মানুষের আস্থাহীনতার কারণ কী? রাশিবিজ্ঞানের সম্ভাব্য অপব্যবহারের কয়েকটি উদাহরণ দাও।

1.4 নীচের সিদ্ধান্তগুলির যাথার্থ্য বিচার কর :

(i) রাশিবিজ্ঞান একটি অত্যন্ত কঠিন বিষয়, কারণ প্রতিবছর যে-সব ছাত্র-

ছাত্রী সাম্মানিক রাশিবিজ্ঞানসহ উত্তীর্ণ হয়, তাদের মধ্যে প্রথম শ্রেণী পায় মাত্র শতকরা 10 জন।

(ii) আকাশবাণীর কলকাতা কেন্দ্রের অনুষ্ঠানসূচী খুবই জনপ্রিয়, কারণ যেসব শ্রোতা এ ব্যাপারে ষ্টেশন-ডিরেক্টরের সঙ্গে পত্রালাপ করেন, তাঁদের প্রায় শতকরা 70 জনই এর প্রশংসা করেন।

(iii) এবারের কলেজ ইউনিয়নের নির্বাচনে একমাত্র মহিলা প্রার্থী ছাত্রীদের শতকরা 85 জনের সমর্থন লাভ করেছেন। সুতরাং ভোটদানে নিঃসন্দেহে পক্ষপাতিত্ব হয়েছে।

(iv) মহিলা-কর্মীরা পুরুষ-কর্মীদের তুলনায় বেশী সময়নিষ্ঠ, কারণ মহাকরণের কর্মীদের মধ্যে শতকরা 80 জন মহিলা এবং শতকরা 45 জন পুরুষ 11টার আগে অফিসে আসেন।

(v) পুরুষদের তুলনায় মেয়েরা ক্যান্সাররোগে কম আক্রান্ত হয়, কারণ গতমাসে চিত্তরঞ্জন ক্যান্সার হাসপাতালে মহিলাদের তিনগুণ পুরুষ-রোগী ভর্তি করা হয়েছে।

(vi) বিছানায় শোয়া খুবই বিপজ্জনক, কারণ আজ পর্যন্ত পৃথিবীতে যত মৃত্যু ঘটেছে তার প্রায় 99 শতাংশ ঘটেছে বিছানাতেই !

(vii) কলকাতার গোয়েন্দা-বিভাগের থেকে দিল্লীর গোয়েন্দা-বিভাগ অনেক বেশী তৎপর, কেননা গতবছর এই দুটি শহরে চুরির আসামী ধরা পড়েছে যথাক্রমে 67টি ও 195টি।

(viii) পশ্চিমবঙ্গে প্রথমশ্রেণীর শহরের সংখ্যা ক্রমশঃ কমার দিকে, কেননা 1961 ও 1971 সালের আদমশুমারিতে এই সংখ্যা ছিল যথাক্রমে 11 ও 5.

## 1.6 নির্দেশিকা

1. Croxton, F. E., and Cowden, D. G. *Applied General Statistics.* Prentice Hall, 1964.

2. Mills, F. C. *Statistical Methods.* H. Holt, 1955.

3. Moroney, M. G. *Facts from Figures.* Penguin, 1956.

4. Wallis, W. A., and Roberts, H. V. *Statistics, a New Approach.* Methuen, 1950.

# 2 রাশিতথ্য আহরণ এবং উপস্থাপন
# ( Collection and Presentation of Statistical Data )

## 2.1 তথ্য আহরণ :

ইতিমধ্যেই ইঙ্গিত দেওয়া হয়েছে যে, রাশিবিজ্ঞানে উপাত্ত বা তথ্যসমষ্টির একটি গুরুত্বপূর্ণ ভূমিকা রয়েছে। প্রকৃতপক্ষে উপাত্তই হ'ল রাশিবিজ্ঞানের মূল আলোচ্য বিষয়বস্তু। সুতরাং এই তথ্যসমষ্টি কোন্ কোন্ সূত্রে এবং কী কী পদ্ধতিতে সাধারণতঃ সংগ্রহ করা হয় তা দিয়ে আমাদের বর্তমান আলোচনা শুরু করা যেতে পারে।

তথ্য **প্রাথমিক সূত্রে** (primary source) অথবা **গৌণ সূত্রে** (secondary source) সংগৃহীত হতে পারে। প্রাথমিক সূত্রে তথ্য-সংগ্রহের তিনটি পদ্ধতি আছে। প্রথমটি হ'ল **প্রত্যক্ষ অবেক্ষণ পদ্ধতি** (direct observation method). এই পদ্ধতিতে সমীক্ষক প্রত্যক্ষভাবে তাঁর প্রয়োজনীয় তথ্য সংগ্রহ ক'রে থাকেন—যেমন কিছুসংখ্যক হৃদ্‌রোগীর রক্তচাপের পরিমাপ নেওয়া, অথবা একটি বিশেষ রাস্তার মোড়ে সকাল নটা থেকে দশটার মধ্যে ক'খানি মোটর-গাড়ী যাতায়াত করল গুণে দেখা। অনেকসময় সমীক্ষকের পক্ষে সরাসরি তথ্য-সংগ্রহ সম্ভব হয় না—যেমন বিভিন্ন পরিবারের মাথাপিছু মাসিক খরচ সম্বন্ধে জানতে হলে পরিবারের কর্তার বিবৃতির উপর নির্ভর করতেই হয়। এইসব ক্ষেত্রে সাধারণতঃ এক বা একাধিক সাক্ষাৎকারী (interviewer) প্রেরণ ক'রে প্রয়োজনীয় তথ্য সংগ্রহ করা হয়, তাই এই পদ্ধতিটিকে বলে **সাক্ষাৎকার পদ্ধতি** (interview method)। সমীক্ষাগত ভৌগোলিক অঞ্চলটি বহুবিস্তৃত হলে অনেক সময় সাক্ষাৎকারী প্রেরণ করাও সম্ভবপর হয় না। সেক্ষেত্রে প্রথমে এমনভাবে একটি **প্রশ্নগুচ্ছ** (questionnaire) রচনা করা হয়, যেন গুচ্ছগত প্রশ্নগুলির উত্তর থেকেই প্রয়োজনীয় তথ্যের সবটুকু পাওয়া সম্ভব হয়। ছাপানো এই প্রশ্নগুচ্ছটি ( সাধারণতঃ এক-একটি প্রশ্নের পাশেই উত্তরের জন্যে ঘর নির্দিষ্ট করা থাকে ) অতঃপর ডাকযোগে বা লোকমারফত পাঠানো হয় সংশ্লিষ্ট বিভিন্ন ব্যক্তি বা প্রতিষ্ঠানের কাছে, উত্তরের জন্যে নির্দিষ্ট ঘরগুলি পূরণ ক'রে পুনরায় এটি সমীক্ষকের কাছে ফেরত পাঠানোর অনুরোধ জানিয়ে। সাধারণতঃ এই

প্রশ্ন-তালিকার সঙ্গে সংশ্লিষ্ট তথ্যগুলি কী উদ্দেশ্যে সংগ্রহ করা হচ্ছে তা ব্যাখ্যা ক'রে উদ্দিষ্ট ব্যক্তি বা প্রতিষ্ঠানকে সহযোগিতা করার অনুরোধ জানানো হয়, এবং সমীক্ষকের ঠিকানা এবং উপযুক্ত ডাকটিকিটসহ একটি খামও পাঠানো হয়। উদাহরণস্বরূপ মাথার যন্ত্রণার একাধিক ওষুধের আপেক্ষিক কার্যকারিতা নির্ণয়ের উদ্দেশ্যে ভারতের বিভিন্ন অঞ্চলের ডাক্তারদের এই ব্যাপারে মতামত সংগ্রহের জন্য এই **প্রশ্নগুচ্ছ-প্রেরণ-পদ্ধতিটি** (questionnaire method) ব্যবহার করা হয়।

অনেক সময় রাষ্ট্র বা কোন প্রতিষ্ঠান বা ব্যক্তি-বিশেষ নিজেদের প্রয়োজনে বা ব্যবসায়িক কারণে নানান বিষয়ের উপর নিয়মিত বিভিন্ন ধরনের তথ্য সংগ্রহ এবং প্রকাশ ক'রে থাকেন। অধিকাংশ ক্ষেত্রেই এই ধরনের কাজ সরকারী উদ্যোগে হয়ে থাকে। তাই এইভাবে সংগৃহীত এবং প্রকাশিত তথ্যকে বলা হয়-**সরকারী পরিসংখ্যান** (Official Statistics)। সমীক্ষক প্রত্যক্ষভাবে তথ্য সংগ্রহের পথে না গিয়ে প্রয়োজনমতো এইসব সরকারী পরিসংখ্যান বা ব্যক্তিগত প্রচেষ্টায় ইতিমধ্যে অন্যত্র প্রকাশিত রাশিতথ্য ব্যবহার করতে পারেন। তথ্য-সংগ্রহের এই সূত্রটি হ'ল পরোক্ষ অথবা গৌণ সূত্র।

সাক্ষাৎকার পদ্ধতিতে তথ্য-সংগ্রহের সব থেকে বড় অসুবিধা হ'ল, এই পদ্ধতিতে সংগৃহীত তথ্য সাক্ষাৎকারীর ব্যক্তিগত পছন্দ-অপছন্দ এবং প্রবণতা দ্বারা প্রভাবিত হওয়ার আশঙ্কা থাকে। সুতরাং এই পদ্ধতিটি ব্যবহার করতে হলে সাক্ষাৎকারীদের উপযুক্ত প্রশিক্ষণ দেওয়ার ব্যবস্থা করতে হয়। প্রশ্নগুচ্ছ-প্রেরণ-পদ্ধতির অসুবিধা হ'ল, ব্যক্তিগত আলস্য, অনিচ্ছা অথবা অন্যান্য নানান কারণে বেশ কিছু সংখ্যক প্রশ্ন-তালিকা সমীক্ষকের কাছে উত্তরসমেত আর ফেরত আসে না। **অনুত্তরের** (non-response) সংখ্যা কমানোর জন্য প্রয়োজনবোধে একই ব্যক্তি বা প্রতিষ্ঠানের কাছে একাধিকবার প্রশ্ন-তালিকা পাঠানোর প্রয়োজন হয়। তাছাড়া, এই পদ্ধতিটি গ্রহণ করলে প্রশ্নগুচ্ছটি রচনা করার সময় খুব সতর্কতা অবলম্বন করতে হয়, যাতে প্রয়োজনীয় কোন তথ্য বাদ না পড়ে, অথবা অপ্রয়োজনীয় কোন তথ্য সংগৃহীত হওয়ার সুযোগ না থাকে এবং প্রশ্নগুলির ভাষা জটিল বা দ্ব্যর্থবোধক না হয়। সরকারী বা অন্য সূত্রে প্রকাশিত পরিসংখ্যান ব্যবহার করার আগে সেগুলির নির্ভরযোগ্যতা সম্বন্ধে নিঃসন্দেহ হওয়া প্রয়োজন।

## 2.2 তথ্য নিরীক্ষণ :

ভুল করা মানুষের স্বাভাবিক ধর্ম। তাই যথেষ্ট সাবধানতা অবলম্বন করা সত্ত্বেও সংগৃহীত তথ্যে কিছু ভুলভ্রান্তি থেকে যাওয়া খুবই সম্ভব। সুতরাং বিশ্লেষণের পূর্বে সংগৃহীত তথ্য থেকে এই ধরনের ভুলভ্রান্তি দূর করার জন্য এগুলির নিরীক্ষণ (scrutiny) একান্ত প্রয়োজন।

নিরীক্ষণের কোন ধরাবাঁধা নিয়ম নাই। নিরীক্ষণের সাফল্য নির্ভর করে নিরীক্ষকের অভিজ্ঞতা, বাস্তববুদ্ধি এবং সাধারণ জ্ঞানের উপর। নিরীক্ষণে সাধারণতঃ ধরা পড়তে পারে এ রকম কয়েক ধরনের ভ্রান্তির কথা আলোচনা করা যেতে পারে।

লিপিবদ্ধ এক-একটি তথ্য প্রথম দৃষ্টিতেই অবাস্তর মনে হতে পারে—স্পষ্টতঃই এগুলি ঘটে অনবধানতাবশতঃ। যেমন মনে কর, বিভিন্ন দিনে কলকাতার সর্বোচ্চ তাপমাত্রা-সংক্রান্ত তথ্য ( ফারেনহাইট ডিগ্রিতে ) পাওয়া গেছে : 96·1, 95·3, 100·4, 1014, 99·5, 9·73, 98·2. এখানে চতুর্থ এবং ষষ্ঠ মান-দুটিতে যে দশমিক বিন্দু-বিভ্রাট ঘটেছে, খুব সহজেই তা বলা যায়।

কোন কোন ক্ষেত্রে বিশেষ একটি তথ্য অসম্ভব না হলেও সহজেই আমাদের সন্দেহ উদ্রেক করতে পারে। যেমন, 7 বৎসর বয়স্কা একটি বালিকাকে যদি বিবাহিতা হিসাবে দেখানো হয়। এইসব ক্ষেত্রে পুনরায় অনুসন্ধান প্রয়োজন।

অনেকসময় আপাতদৃষ্টিতে ভ্রমশূন্য মনে হলেও কোন ব্যক্তি-সংক্রান্ত সংগৃহীত একাধিক তথ্য পরস্পর বিরোধী হতে পারে—যেমন কোন ব্যক্তির ঘোষিত জন্ম-তারিখ এবং বয়সের মধ্যে অসামঞ্জস্য থাকা সম্ভব। এইসব ক্ষেত্রেও কোন্ তথ্যটি সঠিক তা জানার জন্য পুনরায় অনুসন্ধান প্রয়োজন।

যোগ, বিয়োগ, গুণ, ভাগ, শতকরা হার—ইত্যাদিতে ভুল থাকা খুবই সম্ভব। সুতরাং সংগৃহীত তথ্যে গাণিতিক পদ্ধতির ব্যবহার থাকলে নিরীক্ষণের সময় সেগুলি ভালোভাবে পরীক্ষা ক'রে নেওয়া দরকার।

## 2.3 তথ্যের প্রকারভেদ :

উপস্থাপন, বিশ্লেষণ, ব্যাখ্যান প্রভৃতি বিভিন্ন স্তরে বিভিন্ন ধরনের তথ্য-সমষ্টির ক্ষেত্রে কিছুটা পদ্ধতিগত বৈসাদৃশ্য হওয়া সম্ভব। তাই শুরুতে তথ্যের প্রকারভেদ নিয়ে সামান্য আলোচনা ক'রে নেওয়া দরকার।

**গুণগত তথ্য** এবং **পরিমাণগত তথ্য** (qualitative and quantitative data) : অনেক সময় সংগৃহীত তথ্য সংখ্যামানের সাহায্যে প্রকাশ করা যায় না। অথচ রাশিবিজ্ঞানসম্মত সমীক্ষায় এগুলি প্রয়োজনে আসে। যেমন, কোন অফিসে কর্মরত সকল কর্মচারী সম্পর্কে তাঁরা স্নাতক কিংবা অস্নাতক এই তথ্য অথবা কোন কারখানায় নির্দিষ্ট আধঘণ্টা পরিমিত সময়ে উৎপন্ন দ্রব্যগুলি ত্রুটিযুক্ত অথবা ত্রুটিমুক্ত এই তথ্য। এগুলি হ'ল গুণগত তথ্যের উদাহরণ।

অধিকাংশ ক্ষেত্রেই সংগৃহীত তথ্য পরিমাণ-নির্দেশক এবং সংখ্যামানে প্রকাশযোগ্য। যেমন, ভারতে বিভিন্ন প্রদেশে শিক্ষিতের হার, বিভিন্ন ব্যক্তির আয়ের পরিমাণ ইত্যাদি। এই ধরনের তথ্যকে বলা হয় পরিমাণগত তথ্য।

**পরিসংখ্যা রাশিতথ্য** এবং **অ-পরিসংখ্যা রাশিতথ্য** (frequancy data and non-frequency data) : তথ্য আহরণের পর অনেক সময় যাদের সম্বন্ধে তথ্য আহরণ করা হ'ল তাদের মধ্যে মোট কতজন বা কতগুলি একই গুণগত তথ্যের আওতায় এলো, বা পরিমাণগত তথ্যের ক্ষেত্রে একটি বিশেষ মানে বা একটি বিশেষ মান-সীমায় পাওয়া গেল মোট কত জনকে বা কতগুলিকে —গণনা ক'রে দেখা হয়, এবং সংগৃহীত তথ্য এই গণনার ফলাফলের আকারে প্রকাশ করা হয়। যেমন, উপরের উদাহরণে বলা যেতে পারে অফিসটিতে 157 জন কর্মচারীর মধ্যে 39 জন স্নাতক এবং 118 জন অস্নাতক। কিংবা যে 1,131 জন ব্যক্তির কাছ থেকে আয়-সংক্রান্ত তথ্য সংগ্রহ করা হ'ল তাদের মধ্যে 451 জনের আয় 100 টাকার নিচে, 326 জনের আয় 101 টাকা থেকে 200 টাকার মধ্যে,.........ইত্যাদি। এই ধরনের দ্বিতীয় পর্যায়ে গনণাসঞ্জাত রাশিতথ্যকে পরিসংখ্যা রাশিতথ্য বলে। লক্ষ্য কর, মূল তথ্যসমষ্টি গুণগত হলেও লব্ধ পরিসংখ্যা রাশিতথ্য সংখ্যায় প্রকাশিত।

রাশিতথ্য এইভাবে দ্বিতীয় পর্যায়ে গণনার ফলাফলসঞ্জাত না হলে আমরা পাই অ-পরিসংখ্যা রাশিতথ্য—যেমন, ভারতের বিভিন্ন প্রদেশে মোট কৃষিজমির পরিমাণ, বিভিন্ন বৎসরে দেশে আগত শরণার্থীদের সংখ্যা। লক্ষ্য কর, দ্বিতীয় উদাহরণে রাশিতথ্যগুলি শুদ্ধসংখ্যা হলেও এগুলি পরিসংখ্যা নয়।

**পরিসংখ্যা** (frequency) কথাটি রাশিবিজ্ঞানে বহুল ব্যবহৃত। এ সম্বন্ধে পরবর্তী পরিচ্ছেদে বিস্তারিত আলোচনা করা হবে।

অ-পরিসংখ্যা রাশিতথ্য আবার বিভিন্ন প্রকৃতির হতে পারে—যেমন, **কালক্রমিক রাশিতথ্য** বা **কালীন সারি** (time series) অথবা

**ভৌগোলিক সারি** (geographical series)। বিশেষ কোন নিয়ম অনুযায়ী লিপিবদ্ধ একপ্রস্থ রাশিকে সংখ্যা-সারি, সারি অথবা রাশিমালা (series) বলা হয়।

কালীন সারির উদাহরণ হ'ল 1951 থেকে 1971 সাল পর্যন্ত বিভিন্ন বৎসরে ভারতে উৎপন্ন গমের পরিমাণ, জুন-জুলাই মাসের বিভিন্ন দিনে কলকাতায় বৃষ্টিপাতের পরিমাণ, ইত্যাদি।

কোন রাশিতথ্য ভৌগোলিক অঞ্চল অনুযায়ী প্রদত্ত হলে আমরা পাই ভৌগোলিক সারি, যেমন ভারতের বিভিন্ন প্রদেশে শিক্ষিত বেকারের সংখ্যা, বিভিন্ন ইউরোপীয় দেশের সঙ্গে ভারতের বহির্বাণিজ্যের পরিমাণ, ইত্যাদি।

## 2.4 রাশিতথ্য উপস্থাপন :

রাশিতথ্য আহরণ এবং নিরীক্ষণের পর বিশ্লেষণের পূর্বে এগুলি পরিচ্ছন্ন এবং সুশৃঙ্খলভাবে সাজানো এবং যাতে সহজে বোধগম্য হয় এমনভাবে পরিবেশন করা প্রয়োজন, কারণ অবিন্যস্ত পর্যায়ে রাশিতথ্যের গুরুত্বপূর্ণ বৈশিষ্ট্যগুলি সহজে চোখে পড়ে না। বর্ণনাত্মক পদ্ধতিতে (by using a paragraph of text), সারণীবিন্যাসের (tabulation) সাহায্যে, এবং লেখ ও চিত্র ব্যবহারযোগে সাধারণতঃ রাশিতথ্য উপস্থাপন করা হয়।

### 2.4.1 বর্ণনাত্মক পদ্ধতি :

এই পদ্ধতিতে এক বা একাধিক অনুচ্ছেদ ব্যবহার ক'রে সংগৃহীত রাশিতথ্য পরিবেশন করা হয়। নিচের উদাহরণটি লক্ষ্য কর।

"পশ্চিমবঙ্গ সরকারের Bureau of Applied Economics and Statistics কর্তৃক প্রকাশিত Report on Earners' Survey 1962 for Calcutta Industrial Areas (excluding Calcutta) নামক পুস্তিকা থেকে সম্প্রতি কলকাতা শিল্পাঞ্চলে (কলকাতা ব্যতীত) বিভিন্ন ধরনের বৃত্তিতে নিযুক্ত শ্রমিকদের মধ্যে বাঙালীদের অনুপাতের একটি শোচনীয় চিত্র পাওয়া গেছে।

এই অঞ্চলের মোট 315·89 হাজার শ্রমিকের মধ্যে বাঙালীদের সংখ্যা মাত্র 99·57 হাজার, অর্থাৎ মোট সংখ্যার মাত্র শতকরা 32 ভাগ। কৃষিকার্য এবং পশুপালন, খনিকার্য, যন্ত্র-সহযোগে নির্মাণকার্য, যন্ত্র-ব্যতিরেকে নির্মাণকার্য (হস্তশিল্প ব্যতীত) এবং অন্যান্য (নির্দিষ্ট)—এই কয়েকটি বৃত্তিতে যথাক্রমে 7·08, 1·06, 163·78, 75·60 এবং 18·34 হাজার জন শ্রমিকের মধ্যে বাঙালীদের সংখ্যা যথাক্রমে 4·24, 0·23, 48·59, 25·74 এবং 3·71 হাজার। তুলনামূলকভাবে, এই কয়টি বৃত্তির মধ্যে একমাত্র কৃষিকার্য ও পশুপালনেই বাঙালী শ্রমিকরা

অর্ধেকের বেশী (60%)। অন্যান্যগুলিতে বাঙালীদের শতকরা হার খুবই শোচনীয় —যথাক্রমে, 22, 30, 34 ও 20. যে সব শ্রমিকদের বৃত্তি নির্দিষ্টভাবে উল্লেখ করা হয়নি তাদের মধ্যে অবাঙালীদের শতকরা হার 66%.

অবশ্য প্রতিবেদনটি থেকে দেখা যাচ্ছে, একমাত্র কৃষিকার্য ও পশুপালন ছাড়া (এই বৃত্তিতে বাঙালী শ্রমিকদের মাথাপিছু মাসিক আয় অবাঙালীদের তুলনায় দু টাকা কম) আর সব বৃত্তিতেই অবাঙালী শ্রমিক অপেক্ষা বাঙালীদের মাথাপিছু আয় বেশী। বিভিন্ন বৃত্তির মধ্যে যন্ত্র-সহযোগে উৎপাদনে নিযুক্ত শ্রমিকদের উপার্জনই সবথেকে ভালো, বাঙালী ও অবাঙালী শ্রমিকদের ক্ষেত্রে প্রতি মাসে যথাক্রমে 95 টাকা ও 76 টাকা। খনিকার্য, যন্ত্র-ব্যতিরেকে নির্মাণ-কার্য (হস্তশিল্প ব্যতীত), অন্যান্য (নির্দিষ্ট) এবং অন্যান্য (অনির্দিষ্ট)—এই বৃত্তিগুলিতে নিযুক্ত অবাঙালী শ্রমিকদের মাথাপিছু মাসিক আয় যথাক্রমে 66, 73, 63 ও 73 টাকা, বাঙালীদের ক্ষেত্রে এই পরিমাণগুলি যথাক্রমে 6, 7, 3 এবং 4 টাকা বেশী। কৃষিকার্য ও পশুপালনে নিযুক্ত বাঙালী শ্রমিকদের মাসিক মাথাপিছু আয় 63 টাকা। বিভিন্ন বৃত্তিতে নিযুক্ত মোট শ্রমিকদের মাথাপিছু মাসিক আয় (টাকায়) হ'ল 64 (কৃষি), 82 (যন্ত্র-সহযোগে নির্মাণকার্য), 67 (খনি), 75 (যন্ত্র-ব্যতিরেকে নির্মাণকার্য), 64 (অন্যান্য—নির্দিষ্ট) এবং 74 টাকা (অন্যান্য—অনির্দিষ্ট)।

সমস্ত বৃত্তির বিচারে শ্রমিকপিছু মাসিক আয় 78 টাকা, বাঙালী ও অবাঙালী শ্রমিকদের ক্ষেত্রে এই পরিমাণ যথাক্রমে 86 টাকা এবং 74 টাকা।"

বর্ণনাত্মক পদ্ধতির সবথেকে বড় অসুবিধা হ'ল, এই পদ্ধতিতে উপস্থাপিত রাশিতথ্য সম্বন্ধে সম্যক ধারণা পেতে হলে পাঠককে বর্ণনাটি সাধারণতঃ একাধিকবার আগাগোড়া পাঠ করতে হয়, যার জন্য অনেক সময় এবং ধৈর্য প্রয়োজন। তাছাড়া এই পদ্ধতিতে সদৃশ তথ্যগুলির তুলনামূলক চিত্রও ভালোভাবে পাওয়া সম্ভব হয় না। তবে প্রদত্ত রাশিতথ্যের বিশেষ বিশেষ অংশে গুরুত্ব আরোপ করার, বা সেইদিকে পাঠকের দৃষ্টি বিশেষভাবে আকৃষ্ট করার প্রয়োজন হলে এই পদ্ধতিতে তা করা সম্ভব।

রাশিতথ্য উপস্থাপনের এই বর্ণনাত্মক পদ্ধতিটি সাধারণতঃ অনুসৃত হয় না।

### 2.4.2 সারণীবিন্যাস :

সারণীর সাহায্যে অপেক্ষাকৃত কার্যকরীভাবে রাশিতথ্য পরিবেশন করা যায়। উপযুক্ত এবং যথাযথভাবে পরিকল্পিত একটি সারণীতে অল্পপরিসরে অনেক বেশী

তথ্য পরিবেশিত হতে পারে। সারণীর সাহায্যে রাশিতথ্য উপস্থাপনের উল্লেখযোগ্য অন্যান্য সুবিধাগুলি হ'ল : সংক্ষিপ্ততা, সহজবোধ্যতা, তুলনামূলক বিচারের সুযোগ, প্রয়োজনমতো তথ্য সহজে অন্যত্র ব্যবহারের সুযোগ, ইত্যাদি। 2.4.1 অনুচ্ছেদে প্রদত্ত রাশিতথ্য পরিবেশনের উদ্দেশ্যে নিম্নলিখিত সারণীটি ব্যবহার করা যেতে পারে :

**সারণী 2.1**

**মাতৃভাষা এবং বৃত্তি অনুযায়ী কলকাতা শিল্পাঞ্চলের ( কলকাতা ব্যতীত ) শ্রমিকদের বিভাজন এবং মাথাপিছু মাসিক গড় আয়**

| বৃত্তি | বাঙালী শ্রমিক | | অবাঙালী শ্রমিক | | মোট শ্রমিক | | বাঙালী শ্রমিকদের শতকরা হার |
|---|---|---|---|---|---|---|---|
| | সংখ্যা ( হাজারে ) | মাথা পিছু গড় আয় ( টাকায় ) | সংখ্যা ( হাজারে ) | মাথা পিছু গড় আয় ( টাকায় ) | সংখ্যা ( হাজারে ) | মাথা পিছু গড় আয় ( টাকায় ) | |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| কৃষিকার্য ও পশুপালন | 4·24 | 63 | 2·84 | 65 | 7·08 | 64 | 60 |
| খনি-কার্য | 0·23 | 72 | 0·83 | 66 | 1·06 | 67 | 22 |
| যন্ত্র-সহযোগে নির্মাণ-কার্য | 48·59 | 95 | 115·19 | 76 | 163·78 | 82 | 30 |
| যন্ত্র-ব্যতিরেকে নির্মাণ-কার্য* | 25·74 | 80 | 49·86 | 73 | 75·60 | 75 | 34 |
| অন্যান্য ( নির্দিষ্ট ) | 3·71 | 66 | 14·63 | 63 | 18·34 | 64 | 20 |
| অন্যান্য** ( অনির্দিষ্ট ) | 17·06 | 77 | 32·97 | 73 | 50·03 | 74 | 34 |
| মোট | 99·57 | 86 | 216·32 | 74 | 315·89 | 78 | 32 |

*হস্তশিল্পকর্মে নিযুক্ত শ্রমিক ব্যতীত। **যাদের অন্যত্র ধরা হয়নি।

উৎস : *Report on Earners' Survey*, 1962 *for Calcutta Industrial Area (excluding Calcutta)* : *Bureau of Applied Economics and Statistics, Govt. of W. Bengal* (1970).

ব্যবহারিক দিক থেকে সারণী **সাধারণ** অথবা **নির্দেশিকা** (general or reference table) এবং **সংক্ষিপ্ত** অথবা **আহৃত** (summarized or derived table)—এই দুই ধরনের হতে পারে। যে সারণীতে কোন একটি বিশেষ প্রসঙ্গে সংগৃহীত সমুদয় তথ্যের সমাবেশ ঘটে সেগুলি হ'ল সাধারণ সারণী। স্বভাবতঃই এগুলি কিছুটা বিস্তারিত। প্রয়োজনবোধে পরবর্তী সময়ে এই প্রসঙ্গে যে কোন তথ্যের জন্য এই সারণীটি নির্দেশ করা হয়, তাই এগুলিকে নির্দেশিকা সারণীও বলা হয়ে থাকে। অনেক সময়ে সংগৃহীত রাশিতথ্যের একটি বিশেষ দিকের উপর আলোচনার জন্য নির্দেশিকা সারণীর সবটাই প্রয়োজন হয় না। সেক্ষেত্রে নির্দেশিকা সারণীর অংশবিশেষ সরাসরি উদ্ধৃত ক'রে, কিংবা সেটির উপর প্রয়োজনীয় কিছু গাণিতিক পদ্ধতি প্রয়োগ ক'রে পাওয়া যায় আহৃত সারণী। একটি নির্দেশিকা সারণী থেকে একাধিক আহৃত সারণী পাওয়া যেতে পারে। নীচের 2.2 সারণীটি একটি আহৃত সারণী।

**সারণী 2.2**

**বৃত্তি অনুযায়ী কলকাতা শিল্পাঞ্চলের ( কলকাতা ব্যতীত )**
**শ্রমিকদের বিভাজন**

| বৃত্তি | শ্রমিকদের সংখ্যা ( হাজারে ) |
|---|---|
| কৃষিকার্য ও পশুপালন | 7·08 |
| খনি-কার্য | 1·06 |
| যন্ত্র-সহযোগে নির্মাণ-কার্য | 163·78 |
| যন্ত্র-ব্যতিরেকে নির্মাণ-কার্য* | 75·60 |
| অন্যান্য | 68·37 |
| মোট | 315·89 |

* হস্তশিল্পে নিযুক্ত শ্রমিক ব্যতীত।

উৎস : 2.1 সারণীর উৎস।

আকারগত বিচারে সারণী **সরল** (simple) অথবা **জটিল** (complex) হতে পারে। একটি সারণীতে একাধিক ধরনের তথ্য সমাবেশ ঘটলে সেটি হয় জটিল, অন্যথায় আমরা পাই সরল সারণী। 2.1 এবং 2.2 সারণী-দুটি যথাক্রমে জটিল এবং সরল। লক্ষণীয়, নির্দেশিকা সারণী যেমন সরল হতে পারে, তেমনি একটি আহৃত সারণীর জটিল হওয়াও খুবই সম্ভব।

সারণী-বিন্যাসের ধরাবাঁধা কোন নিয়ম নেই। তবে রাশিতথ্য উপস্থাপনের এই পদ্ধতিটি কার্যকরী করার জন্য এই প্রসঙ্গে নিম্নলিখিত বিষয়গুলির দিকে লক্ষ্য রাখতে হবে।

সারণীর একটি উপযুক্ত এবং স্বয়ংব্যাখ্যাত শিরোনামা দেওয়া প্রয়োজন, যাতে এই শিরোনামা থেকেই পাঠক সারণীতে পরিবেশিত রাশিতথ্যের প্রকৃতি সম্বন্ধে সঠিক ধারণা পেতে পারে। ভবিষ্যৎ নির্দেশনার প্রয়োজনে সারণীটির একটি ক্রমিকসংখ্যা দেওয়াও প্রয়োজন।

সারণীর বিভিন্ন স্তম্ভে কী কী রাশিতথ্য উপস্থাপিত হয়েছে তার বর্ণনা দিতে হবে উপযুক্ত শীর্ষ এবং উপশীর্ষ ব্যবহারের সাহায্যে। প্রথম স্তম্ভটি সাধারণতঃ ব্যবহার করা হয় বিভিন্ন সারিতে কী কী রাশিতথ্য পরিবেশিত হয়েছে তার বর্ণনা দেওয়ার উদ্দেশ্যে। স্তম্ভগুলিও ক্রমিকসংখ্যাযুক্ত হওয়া উচিত।

সাধারণতঃ যে সব তথ্যের মধ্যে তুলনার প্রয়োজন, সারণীতে সেগুলি যথাসম্ভব পাশাপাশি রাখা সুবিধাজনক। সারণীতে প্রদত্ত বিভিন্ন ধরনের রাশি-তথ্যের গুরুত্ব অনুযায়ী স্তম্ভগুলির পরিসরবণ্টনে তারতম্য করা এবং স্তম্ভরচনায় বিভিন্ন প্রকার ( স্থূল এবং সূক্ষ্ম ) রেখা ব্যবহার করা যেতে পারে।

বিভিন্ন স্তম্ভে পরিবেশিত রাশিতথ্যের মাপনা একক (unit of measurement) উল্লেখ করা একান্ত প্রয়োজন। সারণীতে প্রদত্ত কোন রাশিতথ্যের জন্য বিশেষ কিছু বক্তব্য থাকলে তা পাদটীকায় দেওয়া হয়। সারণীর শেষে সারণীতে প্রদত্ত রাশিতথ্যের উৎস উল্লেখ করা একটি প্রচলিত প্রথা।

### 2.4.3 লৈখিক পদ্ধতি ঃ

রাশিতথ্য উপস্থাপনায় লেখচিত্রের ব্যবহার একটি বহুলপ্রচলিত পদ্ধতি। বিশেষ ক'রে কালীন সারির ক্ষেত্রে এই পদ্ধতিটি খুবই উপযোগী। এই পদ্ধতিতে সাধারণতঃ **রেখাচিত্র** (line diagram) ব্যবহার করা হয়ে থাকে। পরস্পর সমকোণে ছেদী দুটি অক্ষরেখা নেওয়া হয়। অনুভূমিক অক্ষটি নেওয়া হয় সময়-

সূচক চলের জন্য, উল্লম্ব অক্ষটি পরিমাণনির্দেশক চলের জন্য। প্রদত্ত সময়সীমাগুলির জন্য প্রাপ্ত সংখ্যামান সময়সীমাগুলির মধ্যবিন্দুর বিপরীতে সুবিধামতো স্কেল ব্যবহারে বিভিন্ন বিন্দুর দ্বারা চিহ্নিত করা হয়। সন্নিহিত বিন্দুগুলি সরলরেখার দ্বারা পরস্পর যুক্ত করা হলে যে চিত্রটি পাওয়া যায় সেইটিই হ'ল রেখাচিত্র। স্পষ্টতঃই কালীন সারিতে বৃদ্ধিহার ধ্রুবক হলে রেখাচিত্রটি একটি সরলরেখায় পর্যবসিত হবে।

একাধিক সমজাতীয় কালীন সারি একই রেখাচিত্রে সন্নিবেশিত ক'রে একটি তুলনামূলক চিত্র পাওয়া যেতে পারে। 2.3 সারণীতে প্রদত্ত ভারতীয় পুরুষ এবং নারীদের প্রত্যাশিত আয়ু-সংক্রান্ত কালীন সারি-দুটি 2.1 চিত্রে উপস্থাপিত হয়েছে।

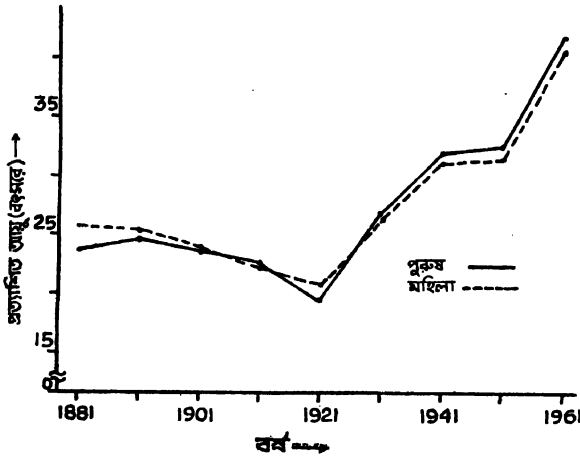
**সারণী 2.3**

**ভারতীয়দের প্রত্যাশিত আয়ু (1881-1961)**

| সাল | প্রত্যাশিত আয়ু ( বৎসরে ) | |
|---|---|---|
| | পুরুষ | নারী |
| (1) | (2) | (3) |
| 1881 | 23·67 | 25·58 |
| 1891 | 24·59 | 25·54 |
| 1901 | 23·63 | 23·96 |
| 1911 | 22·59 | 23·31 |
| 1921 | 19·42 | 20·91 |
| 1931 | 26·91 | 26·56 |
| 1941 | 32·09 | 31·37 |
| 1951 | 32·45 | 31·66 |
| 1961 | 41·89 | 40·55 |

উৎস : *Statistical Abstract of India*, 1968

অনেক সময় একটি সাধারণ অনুভূমিক অক্ষরেখার সঙ্গে একাধিক উল্লম্ব অক্ষরেখার ব্যবহারে রেখাচিত্র সহযোগে পরস্পর সম্পর্কযুক্ত একাধিক কালীন সারির অন্তর্নিহিত সম্পর্কটি পরিস্ফুট করা যায়। এই ধরনের রেখাচিত্রকে বলা

হয় **বহু-অক্ষ** (multiple-axis) রেখাচিত্র। 2.2 চিত্রে দ্বি-অক্ষ রেখাচিত্রের সাহায্যে 1950—1968 সালের জন্য ভারতীয় রেলে যাত্রিসংখ্যা এবং যাত্রিপথের

চিত্র 2.1

ভারতীয়দের ( পুরুষ ও মহিলা ) প্রত্যাশিত আয়ুর রেখাচিত্র।

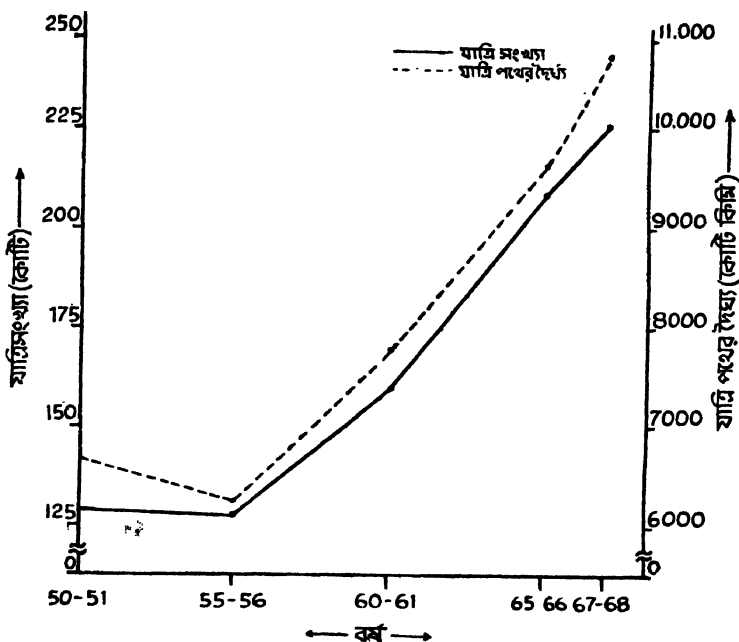
দৈর্ঘ্য—এই দুটি কালীন সারির সম্পর্ক দেখানো হয়েছে। লক্ষ্য কর, এই চিত্রে ব্যবহৃত উল্লম্ব অক্ষ-দুটির স্কেল এবং মাপনা-একক ভিন্ন।

## সারণী 2.4

## ভারতীয় রেলে যাত্রিসংখ্যা এবং যাত্রিপথের দৈর্ঘ্য (1950-51 থেকে 1967-68)

| সাল | যাত্রিসংখ্যা ( কোটি ) | যাত্রিপথের দৈর্ঘ্য ( কোটি কি.মি. ) |
|---|---|---|
| (1) | (2) | (3) |
| 1950-51 | 128·4 | 6651·7 |
| 1955-56 | 127·5 | 6240·0 |
| 1960-61 | 159·4 | 7766·5 |
| 1965-66 | 208·2 | 9629·4 |
| 1967-68 | 225·7 | 10716·3 |

উৎস : *Indian Railways*, 1967-68.

প্রদত্ত কালীন সারি যদি একাধিক উপাদান-সমন্বয়ে গঠিত হয় সেক্ষেত্রে **বন্ধনীচিত্রের** (band chart or component parts chart) ব্যবহারে সময়ের অগ্রগতির সঙ্গে সঙ্গে মোট পরিমাণের বিভিন্ন উপাদানের তুলনামূলক অবদানের একটি সঠিক চিত্র দেওয়া সম্ভব। এক্ষেত্রে মোট পরিমাণে বিভিন্ন উপাদানের

চিত্র 2.2

রেখাচিত্রে ভারতীয় রেলে ( 1950-51 থেকে 1967-68 )
যাত্রিসংখ্যা ও যাত্রিপথের দৈর্ঘ্য।

অবদান পৃথক্ পৃথক্ বিন্দুর সাহায্যে দেখানো হয়—অর্থাৎ এক-একটি সময়-বিন্দুর বিপরীতে নেওয়া হয় একাধিক বিন্দু। এক-একটি উপাদানের জন্য গৃহীত বিন্দুগুলির সাহায্যে এক-একটি রেখাচিত্র পাওয়া যায়। এইভাবে মোট পরিমাণ-সূচক রেখাচিত্রটি একাধিক বন্ধনীতে বিভক্ত হয়ে পড়ে—সেই কারণেই বন্ধনীচিত্র নামটি। বোঝানোর সুবিধার জন্য বিভিন্ন বন্ধনীগুলি বিভিন্নভাবে চিত্রিত হতে পারে। 2.5 সারণীতে প্রদত্ত রাশিতথ্য নীচে একটি বন্ধনীচিত্রের সাহায্যে পরিবেশিত হয়েছে ( চিত্র 2.3 )।

## সারণী 2.5

### ভারতে খাদ্যশস্য উৎপাদন (1950-51 থেকে 1967-68)

| সাল | উৎপাদন ( লক্ষ কুইণ্টালে ) | | | |
|---|---|---|---|---|
| | চাল | গম | অন্যান্য | মোট |
| (1) | (2) | (3) | (4) | (5) |
| 1950-51 | 20·6 | 6·5 | 23·7 | 50·8 |
| 1955-56 | 27·6 | 8·8 | 30·5 | 66·9 |
| 1960-61 | 34·6 | 11·0 | 36·4 | 82·0 |
| 1961-62 | 35·7 | 12·1 | 34·9 | 82·7 |
| 1962-63 | 33·2 | 10·8 | 36·2 | 80·2 |
| 1963-64 | 37·0 | 9·9 | 33·7 | 80·6 |
| 1964-65 | 39·0 | 12·3 | 37·7 | 89·0 |
| 1965-66 | 30·7 | 10·4 | 30·9 | 72·0 |
| 1966-67 | 30·4 | 11·4 | 32·4 | 74·2 |
| 1967-68 | 37·9 | 16·4 | 41·1 | 95·6 |

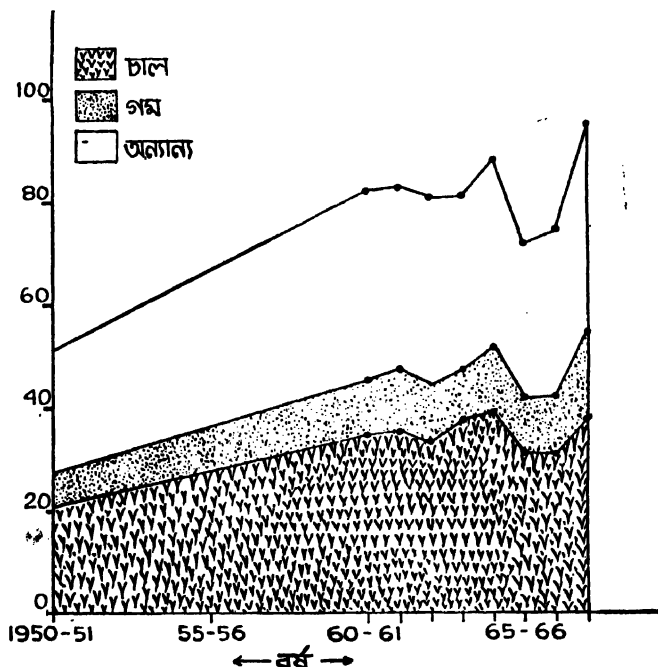
উৎস : *Statistical Abstract of India*, 1968.

উল্লম্ব অক্ষরেখাটি অনেক সময় সাধারণ স্কেলে চিহ্নিত করার পরিবর্তে লগ-স্কেলে চিহ্নিত করা হয়—অর্থাৎ, এই অক্ষরেখায় সমপরিমাণ দৈর্ঘ্য সমান অনুপাত সূচিত করে। এই ধরনের রেখাচিত্রকে **একাক্ষ লগ-চিত্র** (semi logarithmic chart) বা **অনুপাত চিত্র** (ratio chart) বলা হয়। অনুরূপভাবে **উভয়াক্ষ লগ-চিত্র** (double logarithmic chart) অঙ্কন করা যেতে পারে। এই ধরনের চিত্রে ব্যবহারের উপযোগী এক বা উভয়দিকে লগ-স্কেলে চিহ্নিত লেখ-কাগজ (graph paper) বাজারে পাওয়া যায়। লক্ষ্য কর, সাধারণ, একাক্ষ লগ এবং উভয়াক্ষ লগ-লেখচিত্রে একটি সরলরেখা যথাক্রমে

$$\left.\begin{array}{ll} y=a+bx & \\ \log y=a+bx, & \text{অর্থাৎ } y=AB^{x} \\ \text{এবং } \log y=a+b\log x, & \text{অর্থাৎ } y=Ax^{b} \end{array}\right\}, \quad \cdots \qquad (2.1)$$

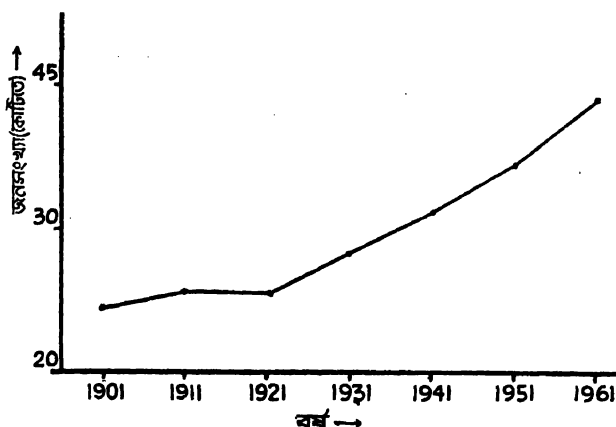
যেখানে $\log A=a$ এবং $\log B=b$,

সমীকরণগুলি নির্দেশ করে। নীচে 2.5 সারণীতে প্রদত্ত ভারতীয় জনসংখ্যা-সংক্রান্ত রাশিতথ্যের একটি অনুপাত চিত্র অঙ্কিত হয়েছে [ চিত্র 2.4 ]।

রেখাচিত্র অঙ্কনের সময় কতকগুলি বিষয়ে লক্ষ্য রাখা প্রয়োজন। প্রথমতঃ চিত্রটির একটি উপযুক্ত শিরোনামা এবং ভবিষ্যৎ নির্দেশনার জন্য একটি ক্রমিক-সংখ্যা দেওয়া দরকার। অনুভূমিক এবং উল্লম্ব অক্ষরেখা-দুটির দৈর্ঘ্য যেন সুসমঞ্জস

চিত্র 2.3

বন্ধনীচিত্রে 1950-51 সাল থেকে 1966-67 সাল পর্যন্ত
ভারতের খাদ্যশস্য উৎপাদনের পরিমাণ।

হয়, সেদিকেও লক্ষ্য রাখতে হবে। অন্যথায় চিত্রটি বিসদৃশ দেখানো ছাড়াও অনুভূমিক রেখার তুলনায় উল্লম্ব রেখা অস্বাভাবিক হ্রস্ব হলে কালীন সারির অন্তর্নিহিত প্রয়োজনীয় চাঞ্চল্য চোখে পড়বে না, অথবা অস্বাভাবিক দীর্ঘ হলে স্বল্পপরিমাণ চাঞ্চল্যও অযথা গুরুত্ব পেয়ে যাবে। অনুভূমিক রেখাটিতে শূন্য (zero) মানটি নির্দেশ করার ব্যাপারটি ঐচ্ছিক—কিন্তু বিভ্রান্তি এড়ানোর জন্য উল্লম্ব রেখায় এটি অবশ্যই নির্দেশ করতে হবে। প্রদত্ত সবকটি মান বেশী বড় হলে নির্বাচিত স্কেলে শূন্য বিন্দুটি দেখানোর জন্য বিন্দুগুলি অনুভূমিক অক্ষ থেকে

অনেক ওপরে অঙ্কিত করতে হতে পারে। এই অসুবিধা দূর করার জন্য উল্লম্ব-রেখায় অনেক সময় একটি সুস্পষ্ট ছেদ ব্যবহার করা হয় [ 2.2 চিত্র ]। আরও

চিত্র 2.4

ভারতের জনসংখ্যার (1901-1961) অনুপাত চিত্র।

লক্ষ্য রাখতে হবে, রেখাচিত্রে যেন পরস্পর তুলনীয় তথ্য উপস্থাপিত হয়। যেমন, কালীন সারিতে গৃহীত সময়সীমাগুলি বিভিন্ন দৈর্ঘ্যসম্পন্ন হলে চিত্রটিতে মোট পরিমাণের পরিবর্তে গড় পরিমাণ-সূচক বিন্দুই সংস্থাপন করা বাঞ্ছনীয়।

### 2.4.4 চিত্রাঙ্কন পদ্ধতি :

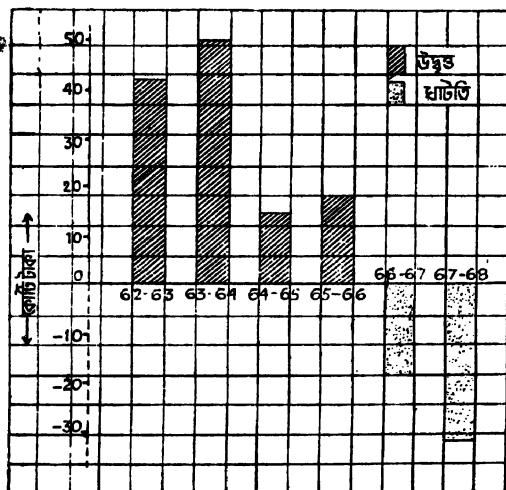
লেখ ব্যতীত অন্য ধরনের চিত্র ব্যবহারেও অনেক সময় রাশিতথ্য উপস্থাপন করা হয়—যেমন, **স্তম্ভচিত্র** (bar-diagram), **রূপচিত্র** (pictogram), **পরিসংখ্যা মানচিত্র** (statistical map), **খণ্ডিত স্তম্ভচিত্র** (divided bar-diagram), **বৃত্তচিত্র** (pie-chart) প্রভৃতি।

**স্তম্ভচিত্র :** একই প্রসারবিশিষ্ট একগুচ্ছ উল্লম্ব কিংবা অনুভূমিক আয়তক্ষেত্র ব্যবহার করা হয় স্তম্ভচিত্রে। সাধারণতঃ কালীন সারির ক্ষেত্রে অনুভূমিক রেখা বরাবর উল্লম্ব আয়তক্ষেত্র এবং ভৌগোলিক ও গুণগত তথ্যের ক্ষেত্রে উল্লম্ব রেখা বরাবর অনুভূমিক আয়তক্ষেত্র নেওয়া হয়ে থাকে। অপর অক্ষটি পরিমাণ-সূচক। লক্ষ্য কর, একই চিত্রে ধনাত্মক ও ঋণাত্মক উভয় শ্রেণীর তথ্যই ( যেমন, লাভ এবং ক্ষতি, উদ্বৃত্ত এবং ঘাটতি, ইত্যাদি ) পরিবেশিত হতে পারে। আয়ত-ক্ষেত্রের দৈর্ঘ্য সংশ্লিষ্ট পরিমাণের সমানুপাতী।

## সারণী 2.6

### ভারতীয় রেলে বাৎসরিক উদ্‌বৃত্ত এবং ঘাটতি
### (1962-63 থেকে 1967-68)

| আর্থিক বৎসর | উদ্‌বৃত্ত (+) বা ঘাটতি (–) (কোটি টাকায়) |
|---|---|
| (1) | (2) |
| 1962-63 | + 42·06 |
| 1963-64 | + 49·24 |
| 1964-65 | + 13·18 |
| 1965-66 | + 18·56 |
| 1966-67 | – 18·27 |
| 1967-68 | – 31·53 |

উৎস : *Indian Railways*, 1967-68.

চিত্র 2.5

ভারতীয় রেলে (1962-63 থেকে 1967-68)
বার্ষিক উদ্‌বৃত্ত ও ঘাটতির স্তম্ভচিত্র।

সন্নিহিত দুটি আয়তক্ষেত্রের মধ্যে একই পরিমাণ ফাঁক (আয়তক্ষেত্রগুলির দৈর্ঘ্যের অর্ধাংশ অপেক্ষা সাধারণতঃ কিছু কম) থাকা প্রয়োজন। কালীন সারির ক্ষেত্রে স্তম্ভগুলির প্রসার সংশ্লিষ্ট সময়সীমার সমানুপাতী হওয়া উচিত। কিন্তু বাস্তবক্ষেত্রে এই নিয়মটি কঠোরভাবে পালন করা সম্ভব হয় না। এক্ষেত্রে উল্লম্ব অক্ষরেখায় শূন্য মানটির প্রদর্শন আবশ্যিক। স্তম্ভগুলি যথাসম্ভব দৈর্ঘ্যের অধঃক্রমানুসারে (বাম থেকে দক্ষিণে অথবা ওপর থেকে নীচে) নেওয়াই প্রথা। কিন্তু কালীন সারির ক্ষেত্রে নিয়মটি মেনে চলা সম্ভব হয় না। এক্ষেত্রে স্তম্ভগুলি সবসময় কালানুক্রমে সাজানো হয়। একই ধরনের একাধিক রাশিতথ্যের তুলনা করার প্রয়োজন হলে **বহু-স্তম্ভ চিত্র** (multiple bar-diagram) ব্যবহার করা হয়। এক্ষেত্রে প্রতিটি সময়সীমা বা ভৌগোলিক অঞ্চল বা গুণগত তথ্যের এক-একটি রূপের জন্য একটির পরিবর্তে একগুচ্ছ আয়তক্ষেত্র ব্যবহার করা হয়।

**সারণী 2.7**

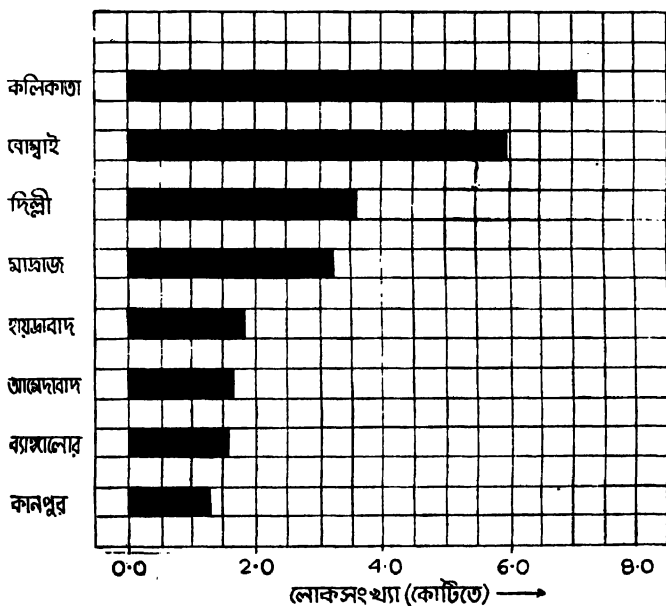
**ভারতের কয়েকটি বড় বড় শহরের জনসংখ্যা**

**(1971 সালের আদমশুমারি অনুযায়ী)**

| শহর * | জনসংখ্যা |
|---|---|
| (1) | (2) |
| কলকাতা | 7,031,832 |
| বোম্বাই | 5,970,575 |
| দিল্লী | 3,647,023 |
| মাদ্রাজ | 3,169,930 |
| হায়দ্রাবাদ | 1,796,339 |
| আমেদাবাদ | 1,741,522 |
| ব্যাঙ্গালোর | 1,653,779 |
| কানপুর | 1,275,242 |
| পুনা | 1,135,034 |

* মূল শহর এবং সন্নিহিত শহরতলী।

উৎস : *The Statesman*, June, 21, 1972.

2.5 চিত্রে একটি কালীন সারি এবং 2.6 চিত্রে একটি ভৌগোলিক সারি উপস্থাপিত হয়েছে স্তম্ভচিত্রের সাহায্যে। 2.7 চিত্রটি একটি বহু-স্তম্ভ চিত্র। এটিতে পরিবেশিত হয়েছে 2.8 সারণীতে প্রদত্ত কিছু গুণগত তথ্য।

চিত্র 2.6

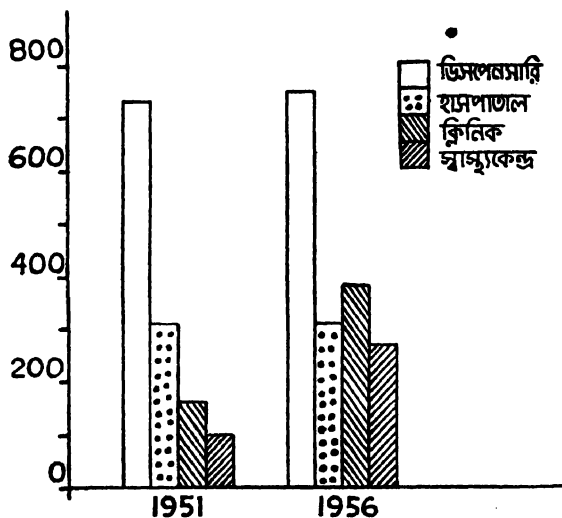
1971 সালের আদমশুমারি অনুযায়ী ভারতের কয়েকটি বড় বড় শহরের জনসংখ্যা।

### সারণী 2.8

### পশ্চিমবঙ্গে* বিভিন্ন ধরনের চিকিৎসাকেন্দ্র

| চিকিৎসাকেন্দ্রের ধরন | সংখ্যা | |
|---|---|---|
| | 1951 | 1956 |
| (1) | (2) | (3) |
| হাসপাতাল | 308 | 308 |
| ডিসপেনসারি | 729 | 749 |
| স্বাস্থ্যকেন্দ্র | 100 | 266 |
| ক্লিনিক | 160 | 379 |
| মোট | 1,297 | 1,702 |

* বিহার থেকে হস্তান্তরিত অঞ্চল বাদে।

উৎস : *Statistical Handbook*, 1970, Govt. of W. Bengal.

চিত্র 2.7

পশ্চিমবঙ্গে বিভিন্ন ধরনের চিকিৎসাকেন্দ্রের সংখ্যার (1951 ও 1956) বহু-স্তম্ভচিত্র।

**রূপচিত্র :** অনেক সময় পরিবেশিত তথ্যের প্রকৃতির সঙ্গে সঙ্গতি রেখে একটি প্রতীক ( সাধারণতঃ একটি ছবি ) বেছে নেওয়া হয় একটি নির্দিষ্ট পরিমাণ রাশিতথ্য সূচিত করার জন্য। ছবিটি বিভিন্ন সংখ্যকবার ( প্রয়োজনবোধে ছবিটির ভগ্নাংশসহ ) ব্যবহার ক'রে রাশিতথ্য উপস্থাপনের এই পদ্ধতিটি হ'ল রূপচিত্র পদ্ধতি।

রূপচিত্রে অনেকে প্রদত্ত সারির বিভিন্ন মান নির্দেশ ক'রে থাকেন গৃহীত ছবিটির বিভিন্ন মাপ ব্যবহার ক'রে। জ্যামিতিক প্রতীক ব্যবহার করা হলে চিত্রটির আয়তনের ভিত্তিতে পরিমাণ নির্দেশ করা হয়।

2.14 সারণীতে প্রদত্ত রাশিতথ্য রূপচিত্রের সাহায্যে অংশতঃ পরিবেশিত হয়েছে 2.8 চিত্রে।

**পরিসংখ্যান মানচিত্র :** একই ভৌগোলিক সীমার বিভিন্ন অঞ্চলের মধ্যে পরিমাণগত তথ্যের তুলনামূলক চিত্র অনেক সময় পরিসংখ্যা-মানচিত্রের

10,000 টোনে

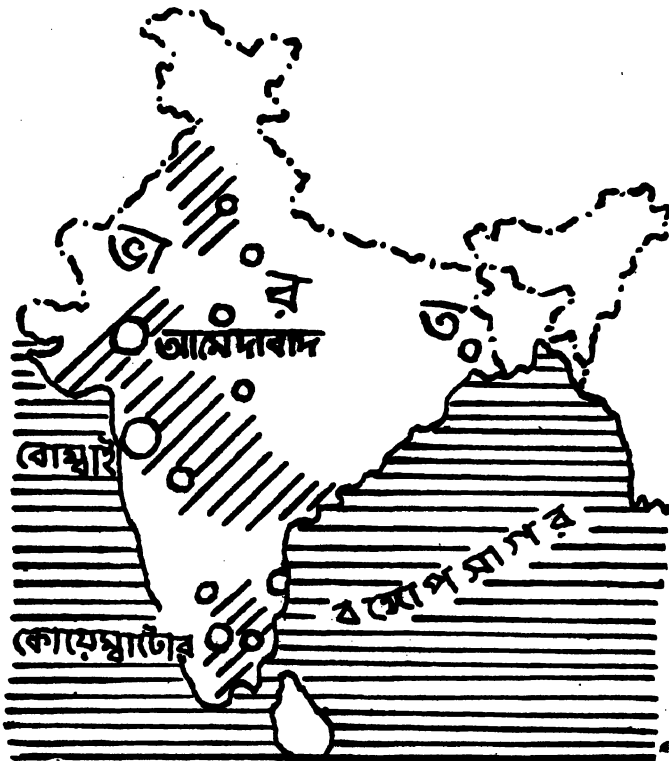
| বর্ষ | কফি উৎপাদনের পরিমাণ (হাজার টোনে) |
|---|---|
| 1951 | 18·4 |
| 1956 | 35·0 |
| 1961 | 67·7 |
| 1966 | 69·0 |

চিত্র 2.8

ভারতের কফি উৎপাদনের রূপচিত্র।

সাহায্যে প্রকাশ করা হয়ে থাকে। জনসংখ্যা, বৃষ্টিপাতের পরিমাণ, বিশেষ একটি কৃষিজাত দ্রব্যের উৎপাদনের পরিমাণ, বিশেষ একটি শিল্পের বিভিন্ন অবস্থিতি, ইত্যাদি ধরনের ভৌগোলিক তথ্য উপস্থাপনের জন্য পরিসংখ্যা-মানচিত্র খুবই উপযোগী। এক্ষেত্রে সমগ্র ভৌগোলিক সীমার একটি মানচিত্র

আঁকা হয় প্রথমে। এই সীমার অন্তর্গত বিভিন্ন অঞ্চলের পরিমাণগত পার্থক্য নির্দেশ করা হয় বিভিন্ন ধরনের চিত্র অথবা শেড (shade), অথবা বিভিন্ন গাঢ়তায় একই শেড ব্যবহার ক'রে। বিভিন্ন সংখ্যক বিন্দুও ব্যবহার করা হয় অনেক সময়। কোন্ ধরনের শেডিং কী সূচনা করছে অথবা একটি বিন্দু কতখানি পরিমাণের নির্দেশক, তার উল্লেখ করা দরকার।

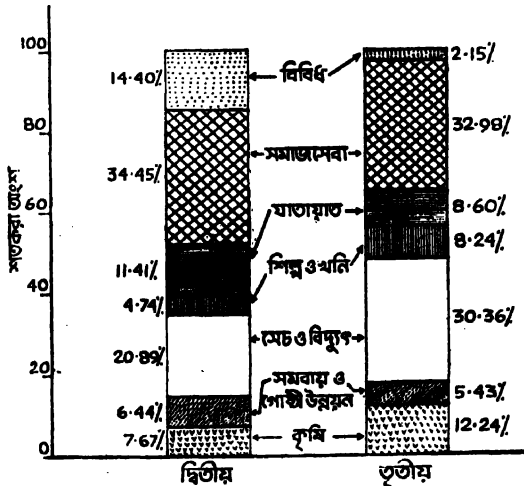
2.9 চিত্রে একটি পরিসংখ্যা-মানচিত্রের সাহায্যে ভারতের কার্পাস-শিল্পের অবস্থিতি দেখানো হয়েছে।

কার্পাস শিল্পকেন্দ্র O

কার্পাস আবাদ ///

চিত্র 2.9

ভারতের কার্পাস-শিল্প।

**খণ্ডিত স্তম্ভচিত্র :** অনেক সময় একটি সমগ্র পরিমাণকে কোন নিরিখ অনুযায়ী কয়েকটি খণ্ডে ভাগ করা হলে মোট পরিমাণে ঐ খণ্ডগুলির আপেক্ষিক অবদান পর্যালোচনা করা প্রয়োজন হতে পারে। সেক্ষেত্রে বিভিন্ন খণ্ডের জন্য রাশিতথ্যগুলি প্রথমে মোট পরিমাণের ভগ্নাংশে কিংবা শতাংশে রূপান্তরিত করা হয়। এই ধরনের রাশিতথ্য চিত্রসহযোগে পরিবেশন করার জন্য খণ্ডিত স্তম্ভচিত্র ব্যবহার করা হয়। উল্লম্ব একটি আয়তাকার স্তম্ভ নেওয়া হয়। এটির আয়তন ( বাস্তবপক্ষে দৈর্ঘ্য ) 1 ( ভগ্নাংশের ক্ষেত্রে ) অথবা 100 ( শতাংশের ক্ষেত্রে ) দ্বারা সূচিত ক'রে বিভিন্ন খণ্ডাংশগুলি নির্দেশ করা হয় স্তম্ভটি নির্দিষ্ট মাপে খণ্ডিত ক'রে। সহজে দৃষ্টিগ্রাহ্য করার জন্য বিভিন্ন খণ্ডগুলিকে বিভিন্নভাবে চিত্রিত করা যেতে পারে। ভূমি থেকে শুরু ক'রে বিভিন্ন খণ্ডগুলি স্থাপন করা হয় যথাসম্ভব মানের অধঃক্রমানুসারে—কিন্তু বিবিধ-তথ্যসূচক (miscellaneous) খণ্ডটি ( যদি থাকে ) সবথেকে শেষে নেওয়াই প্রথা।

সমজাতীয় একাধিক রাশিতথ্যের এই ধরনের খণ্ড অনুপাতের একটি তুলনামূলক চিত্র পাওয়ার জন্যও খণ্ডিত স্তম্ভচিত্র ব্যবহার করা হয়ে থাকে। 2.10 চিত্রে দুটি খণ্ডিত স্তম্ভচিত্রের সাহায্যে দ্বিতীয় ও তৃতীয় পঞ্চবার্ষিকী পরিকল্পনায় পশ্চিমবঙ্গ রাজ্যে বিভিন্ন খাতে ব্যয়ের ( সারণী 2.9 ) তুলনামূলক চিত্র দেওয়া হয়েছে :

চিত্র 2.10

**দ্বিতীয় ও তৃতীয় পঞ্চবার্ষিকী পরিকল্পনায় পশ্চিমবঙ্গে বিভিন্নখাতে ব্যয়ের তুলনামূলক চিত্র।**

**সারণী 2.9**

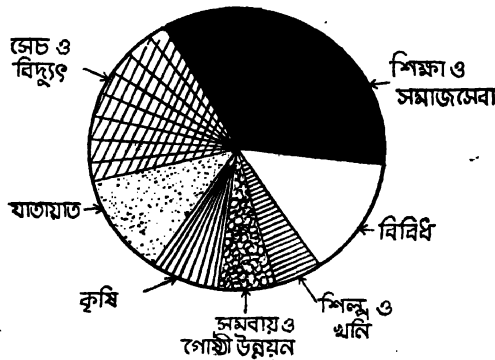
**দ্বিতীয় ও তৃতীয় পঞ্চবার্ষিকী পরিকল্পনায় পশ্চিমবঙ্গ রাজ্য খাতে ব্যয়**

| বিবরণ | ব্যয় ( লক্ষ টাকায় ) | |
|---|---|---|
| | দ্বিতীয় পরিকল্পনা | তৃতীয় পরিকল্পনা |
| (1) | (2) | (3) |
| কৃষি | 1,140·04 | 3,730·05 |
| সমবায় ও গোষ্ঠী-উন্নয়ন | 9, 55·71 | 1,655·98 |
| সেচ ও বিদ্যুৎ | 3,103·85 | 9,251·99 |
| শিল্প ও খনি | 703·90 | 2,511·62 |
| যাতায়াত | 1,695·30 | 2,619·91 |
| সমাজসেবা | 5,118·06 | 10,049·39 |
| বিবিধ | 2,139·79 | 655·58 |
| মোট | 14,856·65 | 30,474·52 |

উৎস : *Statistical Hand Book* (1970), Govt. of West. Bengal.

**বৃত্তচিত্র :** চিত্র-সহযোগে ভগ্নাংশ এবং শতাংশস্থচক রাশিতথ্য পরিবেশনের আর একটি পদ্ধতি হ'ল বৃত্তচিত্র অঙ্কন। একটি বিন্দুর চতুঃপার্শ্বস্থ কোণের পরিমাণ হচ্ছে 360°—স্থতরাং ভগ্নাংশ বা শতাংশগুলিকে 360° ডিগ্রির ভগ্নাংশে রূপান্তরিত ক'রে ( অর্থাৎ শতাংশগুলিকে 3·6 দ্বারা গুণ ক'রে ) একটি বৃত্তের কেন্দ্রে অঙ্কিত অনুরূপ পরিমাণ কোণের সাহায্যেও বিভিন্ন খণ্ডগুলি নির্দেশ করা চলে। এইভাবে প্রাপ্ত চিত্রটির নাম বৃত্তচিত্র। উল্লম্ব ব্যাসার্দ্ধ থেকে শুরু ক'রে ঘড়ির কাঁটার গতিপথ অনুযায়ী যথাসম্ভব মানের অধঃক্রম অনুযায়ী কোণগুলি পর পর আঁকা হয়—অবশ্য যথারীতি বিবিধ তথ্যস্থচক কোণটি আসে সবার শেষে। লক্ষ্য কর, বৃত্তচিত্রে উপস্থাপিত ভগ্নাংশ বা শতাংশগুলি বৃত্তাংশগুলির ক্ষেত্রফলের বা বৃত্তচাপগুলির দৈর্ঘ্যের বা কেন্দ্রস্থ কোণগুলির পরিমাণের সমানুপাতী।

2.10 চিত্রের প্রথম স্তম্ভচিত্রটির বৃত্তচিত্র অঙ্কিত হয়েছে 2.11 চিত্রে।

বৃত্তচিত্র এবং খণ্ডিত স্তম্ভচিত্র একই উদ্দেশ্যে ব্যবহৃত হলেও সাধারণ মানুষের চোখে প্রথমোক্তটির আবেদন বেশী। বিশেষতঃ অর্ধাংশ, চতুর্থাংশ ইত্যাদির ধারণা খণ্ডিত স্তম্ভচিত্রের তুলনায় বৃত্তচিত্রে স্পষ্টতর।

চিত্র 2.11

দ্বিতীয় পঞ্চবার্ষিকী পরিকল্পনায় পশ্চিমবঙ্গ রাজ্যখাতে ব্যয়ের বৃত্তচিত্র।

রাশিতথ্য উপস্থাপনে চিত্র এবং লেখ ব্যবহার সংক্রান্ত বিস্তারিত আলোচনা করা হ'ল। রাশিতথ্য উপস্থাপনের এই পদ্ধতিটির স্বপক্ষে সবথেকে বড় যুক্তি হ'ল, সাধারণ মানুষের কাছে, বিশেষতঃ নিরক্ষরদের কাছে, এটির আবেদন অপরিসীম। তবে ছবি থেকে পরিবেশিত রাশিতথ্য সম্বন্ধে কেবল একটা মোটামুটি ধারণা পাওয়া যায় মাত্র—প্রকৃত মান সম্বন্ধে সঠিক ধারণা পাওয়া সম্ভব হয় না। তাছাড়া পদ্ধতিটি সাধারণতঃ অধিকতর শ্রম এবং ব্যয়-সাপেক্ষ।

বাস্তবক্ষেত্রে সাধারণতঃ রাশিতথ্য পরিবেশনে সারণীর পরিপূরক হিসাবে উপযুক্ত লেখ বা চিত্র ব্যবহার করা হয়।

এখানে অ-পরিসংখ্যা রাশিতথ্য উপস্থাপনার কথা বলা হ'ল। পরিসংখ্যা-রাশিতথ্য উপস্থাপনার প্রসঙ্গটি পরবর্তী অধ্যায়ে বিস্তারিতভাবে আলোচিত হবে।

## 2.5 অনুশীলনী

**2.1** উপাত্ত বা তথ্য আহরণের বিভিন্ন পদ্ধতিগুলির তুলনামূলক আলোচনা কর। কী কী ধরনের উপাত্ত হতে পারে ?

**2.2** রাশিতথ্য উপস্থাপনে সারণী ব্যবহারের উপযোগিতা বর্ণনা কর। কী

কী ধরনের সারণী হতে পারে উদাহরণ সহযোগে আলোচনা কর। সারণী-বিন্যাসের সময় কোন্ কোন্ বিষয়ে লক্ষ্য রাখা উচিত ?

2.3 লেখ এবং চিত্র ব্যবহারে রাশিতথ্য উপস্থাপনের বিভিন্ন পদ্ধতিগুলি বর্ণনা কর। এই উপায়ে কালীন সারি উপস্থাপনের বিভিন্ন পদ্ধতিগুলির তুলনামূলক আলোচনা কর।

2.4 বিভিন্ন ধরনের লেখচিত্রের বর্ণনা দাও। রেখাচিত্র অঙ্কনের সময় কোন্ কোন্ বিষয়ে লক্ষ্য রাখা উচিত ?

2.5 লেখ ও চিত্র ব্যবহার-যোগে ভৌগোলিক রাশিতথ্য পরিবেশনের বিভিন্ন পদ্ধতি বর্ণনা কর।

2.6 স্তম্ভচিত্রের সাহায্যে সাধারণতঃ কী কী ধরনের রাশিতথ্য পরিবেশিত হতে পারে ? উদাহরণ সহযোগে বর্ণনা কর।

2.7 জনৈক সমীক্ষক ভারতের একটি জেলা সম্বন্ধে নিম্নলিখিত তথ্যগুলি সংগ্রহ করেছেন। এগুলি নিরীক্ষণ ক'রে সন্দেহজনক তথ্যগুলি পৃথক কর এবং এগুলির যাথার্থ্য সম্বন্ধে তোমার সন্দেহের কারণ বিবৃত কর :

(i) মোট আয়তন ... ... 2,536 বর্গমাইল
(ii) মোট থানার সংখ্যা ... ... 10
(iii) প্রতি বর্গমাইলে জনসংখ্যার ঘনত্ব ... 252
(iv) মোট পরিবারের সংখ্যা ... ... 9653
(v) মোট চাষের জমি ... ... 1395 বর্গমাইল
(vi) বছরে একবারের বেশী চাষ হয় এমন জমি 232·2 "
(vii) মোট বনভূমি ... ... 601·7 "
(viii) জলভাগের আয়তন ... ... 638·3 "
(ix) থানার গড় আয়তন ... ... 250·0 "
(x) পরিবারের গড় সদস্যসংখ্যা ... ... 5·9
(xi) মাসে গড় চাল খরচ ... ... 75232·2 মণ
(xii) হিন্দু, মুসলমান এবং অন্যান্য সম্প্রদায়ের শতকরা হার ... যথাক্রমে 63·3, 36·2 ও 1·9

2.8 কলকাতা ও বোম্বাই বন্দরের মাধ্যমে 1971 ও 1972 সালে খাদ্যশস্য, অর্থকরী কৃষিপণ্য ও শিল্পজাত দ্রব্যের আমদানী ও রপ্তানী বাণিজ্যের মোট মূল্য-সংক্রান্ত রাশিতথ্য পরিবেশনের জন্য একটি আদর্শ সারণীর ছক প্রস্তুত কর।

সারণীটিতে বিভিন্ন স্তম্ভ ও উপস্তম্ভের যোগফল এবং 1971 সালের তুলনায় 1972 সালে বাণিজ্যের পরিমাণের শতকরা হ্রাস বা বৃদ্ধি প্রদর্শনের ব্যবস্থা থাকা প্রয়োজন।

2.9 নিম্নলিখিত অনুচ্ছেদগুলিতে প্রদত্ত রাশিতথ্য উপযুক্ত সারণীর সাহায্যে পরিবেশন কর :

(i) ইছাপুর পাবলিক লাইব্রেরীর বার্ষিক বিবরণী (1970) থেকে পাঠাগারটির সভ্য-সভ্যাদের পাঠাভ্যাস সংক্রান্ত নিম্নলিখিত তথ্য পাওয়া গেছে :

"জুন (1970) মাসে ধার দেওয়া মোট 3,713 খানি পুস্তকের মধ্যে 2,100 খানি গল্প-উপন্যাস ছিল। পাঠাগারে মোট পাঁচ ধরনের সভ্য আছে—*A*, *B*, *C*, *D* ও *E*. মোট 467 জন সভ্যের মধ্যে প্রথম চারটি বিভাগের সভ্যসংখ্যা যথাক্রমে 15, 176, 98 এবং 129 এবং আলোচ্য এইসব সভ্যদেরকে দেওয়া গল্প-উপন্যাসের সংখ্যা যথাক্রমে 103, 1187, 647 এবং 58. পাঠ্য-পুস্তক এবং গল্প-উপন্যাস ব্যতীত অন্যান্য পুস্তক এই কয়টি বিভাগের সভ্যদের দেওয়া হয়েছিল যথাক্রমে 4, 390, 217 এবং 341 খানি। পাঠ্য-পুস্তক নিয়েছেন কেবল *C*, *D* এবং E জাতীয় সভ্যরা যথাক্রমে 3, 317 এবং 160 খানি।

আলোচ্যমাসে ধার দেওয়া মোট 1,246 খানি পত্র-পত্রিকার মধ্যে মাত্র 396 খানি ছিল টেকনিক্যাল এবং এগুলি মাত্র *B* ( 36 খানি ), *D* ( 45 খানি ) এবং *E* ( 315 খানি ) শ্রেণীর সভ্যদের দেওয়া হয়েছিল। অন্যান্য পত্র-পত্রিকা *B*, *C*, *D* ও *E* শ্রেণীর সভ্যরা নিয়েছিলেন যথাক্রমে 419, 26, 231 এবং 99 খানি।

বিগত মাসের তুলনায় বই দেওয়ার সংখ্যা 3·9% বৃদ্ধি পেলেও পত্র-পত্রিকার ক্ষেত্রে হ্রাস পেয়েছে 6·1%।

(ii) 1971 সালের নভেম্বর মাসের কোন এক দিনের অমৃতবাজার পত্রিকা থেকে পাওয়া খবর :

"গত 18 বছরের সঙ্গে সঙ্গতি রেখে এয়ার ইণ্ডিয়া 1970-71 সালেও ভাল লাভই করেছে। আলোচ্য বছরে এই লাভের পরিমাণ 4·58 কোটি টাকা, 1969-70 সালের তুলনায় যা 28 লাখ টাকা বেশী।

প্রচণ্ড অর্থনৈতিক প্রতিকূল অবস্থার মধ্যেও সংস্থাটির বিমান চালনার সংখ্যা গত বছরের তুলনায় আলোচ্য বছরে 7·2% বৃদ্ধি পেয়েছে। বিমান পরিবহণ-শিল্পে সাম্প্রতিক সমস্যাসঙ্কুল পটভূমির পরিপ্রেক্ষিতে এই হার খুবই সন্তোষজনক।

গত বছরের তুলনায় বর্তমান বছরে সংস্থাটির মালবহনের মোট সুযোগ ( লক্ষ টোনে-কিলোমিটারে ) 506·11 থেকে বৃদ্ধি পেয়ে দাঁড়িয়েছে 515·53-এ, এবং এই দুবছরে প্রকৃতপক্ষে এই সুযোগ ব্যবহৃত হয়েছে ( লক্ষ টোনে-কিলোমিটারে ) যথাক্রমে 256·57 ও 275·17, অর্থাৎ ব্যবহৃত সুযোগের শতকরা হার 1969-70 সালের তুলনায় আলোচ্য বছরে 50·7 থেকে বৃদ্ধি পেয়ে 54·3-এ দাঁড়িয়েছে।

2.10 'নিরালা' বন্ধুগোষ্ঠীর 1972 সালে বিভিন্ন মাসের ব্যয়-সংক্রান্ত নিম্নলিখিত তথ্য উপযুক্ত লেখ বা চিত্রের সাহায্যে উপস্থাপিত কর :

| মাস | ব্যয় ( টাকায় ) | মাস | ব্যয় ( টাকায় ) |
|---|---|---|---|
| জানুয়ারি | 739·27 | জুলাই | 763·05 |
| ফেব্রুয়ারি | 683·74 | অগস্ট | 702·21 |
| মার্চ | 786·15 | সেপ্টেম্বর | 741·11 |
| এপ্রিল | 787·33 | অক্টোবর | 632·82 |
| মে | 743·13 | নভেম্বর | 755·39 |
| জুন | 712·92 | ডিসেম্বর | 737·04 |

2.11 পশ্চিমবঙ্গ সরকারের স্বাস্থ্যখাতে ব্যয়সংক্রান্ত কিছু তথ্য নীচে পরিবেশিত হয়েছে। রেখাচিত্রের সাহায্যে স্বাস্থ্যখাতে মোট ব্যয় ( মোট শুল্কের শতকরা হার ) এবং অনুপাতচিত্র ও রূপচিত্রের সাহায্যে স্বাস্থ্যখাতে মাথাপিছু ব্যয়-সংক্রান্ত তথ্য উপস্থাপিত কর।

একটি বহু-অক্ষ রেখাচিত্রে এই সারণীতে প্রদত্ত সমগ্র তথ্য পরিবেশন কর।

### সারণী 2.10
### পশ্চিমবঙ্গ সরকার কর্তৃক স্বাস্থ্যখাতে ব্যয়

| সাল | স্বাস্থ্যখাতে ব্যয় | |
|---|---|---|
| | মোট ( মোট শুল্কের শতকরা হার ) | মাথাপিছু (টাকায়) |
| 1951-52 | 13·3 | 1·82 |
| 52-53 | 15·0 | 2·18 |
| 53-54 | 14·9 | 2·32 |
| 54-55 | 16·8 | 2·62 |
| 55-56 | 15·4 | 2·78 |
| 56-57 | 16·7 | 3·16 |
| 57-58 | 15·7 | 3·40 |
| 58-59 | 12·2 | 2·95 |
| 59-60 | 14·5 | 3·95 |
| 60-61 | 15·9 | 4·72 |
| 61-62 | 12·0 | 3·25 |

উৎস : *Health on the March*, 1948-61, *West Bengal.*

2.12 উপযুক্ত লেখ ব্যবহারে নিম্নলিখিত সারণীতে প্রদত্ত কালীন সারিগুলির একটি তুলনামূলক চিত্র দাও :

### সারণী 2.11
### ভারতে শিল্পোৎপাদনের সূচক-সংখ্যা
### ( ভিত্তিকাল 1960=100 )

| সাল | সূচক সংখ্যা | | |
|---|---|---|---|
| | সাধারণ | বিদ্যুৎ শিল্প | খনিজ শিল্প |
| 1961 | 109·2 | 110·0 | 105·4 |
| 1962 | 119·7 | 130·3 | 115·2 |
| 1963 | 129·7 | 153·0 | 123·2 |
| 1964 | 140·9 | 174·2 | 119·4 |
| 1965 | 150·9 | 204·4 | 131·7 |
| 1966 | 152·4 | 224·9 | 137·1 |
| 1967 | 150·7 | 243·4 | 133·1 |

উৎস : *Statistical Abstract of India*, 1968.

2.13

**সারণী 2.12**

**পৃথিবীর প্রধান প্রধান রবার উৎপাদনকারী দেশগুলিতে 1968 সালে রবার উৎপাদনের পরিমাণ**

| দেশ | রবার উৎপাদনের পরিমাণ ( হাজার টোনে ) |
|---|---|
| মালয়েশিয়া | 1,042 |
| ইন্দোনেশিয়া | 740 |
| থাইল্যাণ্ড | 254 |
| শ্রীলঙ্কা | 146 |
| ভারত | 68 |
| কম্বোডিয়া | 51 |
| অন্যান্য | 294 |
| মোট | 2,595 |

উৎস : *The Times of India Year Book*, 1970.

উপযুক্ত চিত্রের সাহায্যে এই রাশিতথ্য পরিবেশন কর এবং মোট উৎপাদনে বিভিন্ন দেশের আপেক্ষিক অবদানের একটি তুলনামূলক চিত্র দাও।

2.14

**সারণী 2.13**

**1961 সালের আদমশুমারির হিসাবে বয়স ও লিঙ্গ অনুযায়ী পশ্চিমবঙ্গের জনসংখ্যা**

| বয়স-গোষ্ঠী | জনসংখ্যা ( লক্ষে ) | | |
|---|---|---|---|
| | পুরুষ | স্ত্রী | মোট |
| 0—4 | 29·68 | 29·97 | 59·65 |
| 5—9 | 23·14 | 23·02 | 46·16 |
| 10—14 | 20·42 | 19·39 | 39·81 |
| 15—19 | 18·12 | 16·19 | 34·31 |
| 20—24 | 16·26 | 13·66 | 29·92 |
| 25—29 | 15·21 | 12·10 | 27·31 |
| 30—34 | 14·12 | 10·69 | 24·81 |
| 35—39 | 12·16 | 8·75 | 20·91 |
| 40—44 | 9·92 | 7·09 | 17·01 |
| 45—49 | 7·97 | 5·87 | 13·84 |
| 50—54 | 6·27 | 4·84 | 11·11 |
| 55—59 | 4·76 | 3·92 | 8·68 |
| 60—64 | 3·40 | 3·02 | 6·42 |
| 65—69 | 2·19 | 2·07 | 4·26 |
| 70 এবং তদূর্ধ্ব | 2·35 | 2·70 | 5·05 |

উৎস : *Census of India, Vol XVI, Part IA, Book II.*

বহুস্তম্ভ চিত্র এবং বয়স-লিঙ্গ-পিরামিডের (age-sex-pyramid) সাহায্যে আলোচ্য বছরে বয়স-অনুযায়ী পশ্চিমবঙ্গের স্ত্রী ও পুরুষ জনসংখ্যার একটি তুলনা-মূলক চিত্র দাও।

[ **টীকা : বয়স-লিঙ্গ-পিরামিড** একটি বিশেষ ধরনের বহু-অক্ষ চিত্র। এখানে একটি সাধারণ উল্লম্ব অক্ষরেখায় বয়ঃসীমাগুলি নেওয়া হয়। অনুভূমিক অক্ষরেখাটি উল্লম্ব অক্ষরেখার উভয়দিকে প্রলম্বিত ক'রে এক দিকটি পুরুষের সংখ্যা

ও অন্যদিকটি স্ত্রীলোকের সংখ্যা সূচিত করার জন্য ব্যবহার করা হয়। এক-একটি বয়ঃসীমার জন্য উভয়দিকে একটি ক'রে অনুভূমিক স্তম্ভ নেওয়া হয়। এইভাবে অঙ্কিত চিত্রটি একটি পিরামিডের রূপ নেয়। ]

2.15 সারণী 2.8-এ প্রদত্ত রাশিতথ্য একটি অনুভূমিক বহুস্তম্ভ চিত্রের সাহায্যে উপস্থাপিত কর।

2.16 সারণী 2.14-তে প্রদত্ত রাশিতথ্য উপযুক্ত লেখচিত্রের সাহায্যে উপস্থাপিত কর।

**সারণী 2.14**

**ভারতে কফি উৎপাদন**

| সাল | কফি উৎপাদনের পরিমাণ ( হাজার টোনে ) |
|---|---|
| 1951 | 18·4 |
| 56 | 35·0 |
| 60 | 51·5 |
| 61 | 65·7 |
| 62 | 47·5 |
| 63 | 58·4 |
| 64 | 67·2 |
| 65 | 60·7 |
| 66 | 69·0 |
| 67 | 75·6 |

উৎস : *Statistical Abstract of India*, 1968.

## 2.6 নির্দেশিকা

1. Croxton, F. E. & Cowden, D. J. *Applied General Statistics*. Prentice Hall, 1964.

2. Goon, A. M., Gupta, M. K., and Dasgupta, B. *Fundamentals of Statistics*. (Vol. I). World Press, 1975.

3. Mills, F. C. *Statistical Methods.* H. Holt & Co., 1955.

4. Moore, P. G. *Principles of Statistical Techniques.* Cambridge University Press, 1969.

5. Myers, J. W. *Statistical Presentation.* Littlefield, Adam & Co., 1956.

# পরিসংখ্যা-বিভাজন
# ( Frequency Distribution ) 3

## 3.1 রাশিতথ্যের সংক্ষেপীকরণ :

গুণগত অথবা পরিমাণগত যাই হোক না কেন, সংগৃহীত রাশিতথ্যের পরিমাণ যদি বিরাট হয়ে দাঁড়ায়, তাহলে অবিন্যস্ত অবস্থায় এই বিপুল পরিমাণ রাশিতথ্যের তাৎপর্য এবং গতি-প্রকৃতি অনুধাবন করা দুঃসাধ্য হয়ে পড়ে। সেক্ষেত্রে যে অবস্থায় তথ্যগুলি সরাসরি সংগৃহীত হয় সে অবস্থা থেকে এগুলিকে প্রথমে প্রয়োজনানুসারে একটু অন্যভাবে কিছুটা সংক্ষেপিতরূপে সাজিয়ে নিতে হবে। এই ধরনের সংক্ষেপীকরণের পথে অবশ্য অনিবার্যভাবে কিছু কিছু তথ্যসার বিসর্জন দিতে হয়, কারণ সংক্ষেপিত রাশিতথ্যে ব্যষ্টিগুলিকে আলাদা আলাদাভাবে চিনে নেওয়ার উপায় থাকে না। তবে এর জন্য আমাদের মূল উদ্দেশ্য সাধারণতঃ ব্যাহত হয় না, কারণ আমরা আগেই বলেছি রাশিবিজ্ঞান সাধারণতঃ কোন সমষ্টির বা সমগ্রকের এক বা একাধিক লক্ষণ সম্বন্ধেই আগ্রহী, ব্যষ্টির নয়। কোন একটি কলেজের ছাত্র-ছাত্রীদের একটি বিশ্ববিদ্যালয় পরীক্ষায় ( পূর্ণমান 100 ) প্রাপ্ত নম্বর সম্বন্ধে তথ্যসংগ্রহ করার উদ্দেশ্য সাধারণতঃ হয়ে থাকে পাশের শতকরা হার, প্রথম বিভাগে উত্তীর্ণদের সংখ্যা, গড় নম্বর ইত্যাদি বিষয়ে অনুসন্ধান। এক্ষেত্রে ছাত্রদের নম্বরগুলি আলাদা আলাদা ভাবে না দিয়ে যদি কতজন 1 থেকে 10-এর মধ্যে, কতজন 11 থেকে 20-এর মধ্যে,........., কতজন 91—100-এর মধ্যে পেয়েছে—এইভাবে সংক্ষেপিত আকারে দেওয়া হয়, তাহলেই আমাদের কাজ চলে যায়। বিশেষ একটি ছাত্রের নম্বর জানার আমাদের কোন প্রয়োজন হয় না।

পরিসংখ্যা-বিভাজন (frequency distribution) নিরূপণের সাহায্যে কি-ভাবে এই ধরনের সংক্ষেপীকরণ সম্ভব সে সম্পর্কে বিস্তারিত আলোচনাই আমাদের বর্তমান পরিচ্ছেদের বিষয়সূচী।

## 3.2 লক্ষণের প্রকারভেদ :

দৈনন্দিন জীবনের নানা প্রয়োজনে অথবা কোন বৈজ্ঞানিক অনুসন্ধানের প্রয়োজনে সাধারণতঃ রাশিতথ্য সংগ্রহ করা হয়ে থাকে কিছুসংখ্যক ব্যষ্টির কোন

না কোন লক্ষণের ওপর। এখন এই লক্ষণ প্রকৃতিভেদে দুরকম হতে পারে—**গুণ-লক্ষণ** (qualitative character or attribute) এবং **পরিমাণসূচক লক্ষণ** বা **চল** (quantitative character or variable)। কিছু কিছু লক্ষণ আছে যেগুলির মাত্রা সংখ্যার সাহায্যে প্রকাশ করা সম্ভব নয়—যেমন, ফুলের রঙ, বই-এর ভাষা, নাগরিকের শিক্ষাগত মান, ইত্যাদি। এগুলিকে বলা হয় গুণ-লক্ষণ। একটি গুণ-লক্ষণের একাধিক রূপ (form) থাকে। যেমন পাঠাগারে রক্ষিত বিভিন্ন পুস্তকের ভাষা—এই গুণ-লক্ষণটির বিভিন্ন রূপ হতে পারে ইংরেজী/বাঙলা/হিন্দি/অন্যান্য ভারতীয় ভাষা/অন্যান্য বিদেশী ভাষা। অথবা, বিভিন্ন ব্যক্তির শিক্ষাগত মান—এই গুণ-লক্ষণটির বিভিন্ন রূপ হতে পারে অশিক্ষিত/প্রাথমিক বিদ্যালয় মানের/মহাবিদ্যালয় ও বিশ্ববিদ্যালয় মানের। লক্ষণীয়, গুণ-লক্ষণের বিভিন্ন রূপ নির্বাচনে কিছুটা ব্যক্তি-নির্ভরতা থাকবেই। সাধারণতঃ উদ্দেশ্য এবং প্রয়োজনের সঙ্গে সঙ্গতি রেখে গুণ-লক্ষণের বিভিন্ন রূপ নির্বাচন করা হয়ে থাকে। যেমন, পুস্তকের ভাষা এই গুণ-লক্ষণটির ওপরে প্রদত্ত রূপগুলি না নিয়ে নেওয়া যেতে পারে—ভারতীয় ভাষা/বিদেশী ভাষা ; কিংবা শিক্ষাগত মানের নির্বাচিত রূপগুলি হতে পারে সাক্ষর/নিরক্ষর।

অন্য এক ধরনের লক্ষণের বিভিন্ন মাত্রা গণনা বা মাপনার সাহায্যে নির্ণয় করা চলে এবং এইসব লক্ষণের উপর সংগৃহীত তথ্য সংখ্যার সাহায্যে প্রকাশ করা যায়—যেমন, ছাত্র-ছাত্রীর পরীক্ষায় প্রাপ্ত নম্বর, বিভিন্ন দেশে শিক্ষিতের হার, কারখানায় উৎপন্ন আলপিনের দৈর্ঘ্য, তুলা-জাত সূতার ভারবহনের ক্ষমতা, দিনের সর্বোচ্চ তাপমাত্রা, পরিবারের সদস্যসংখ্যা, একটি বিশেষ মাসে পাঠাগারের পাঠকগণ কর্তৃক ধার নেওয়া পুস্তকের সংখ্যা, ইত্যাদি। গণনযোগ্য বা মাপন-যোগ্য এইসব লক্ষণকে বলা হয় পরিমাণসূচক লক্ষণ বা চল।

কোন গুণ-লক্ষণের ওপর সংগৃহীত তথ্য হ'ল গুণগত উপাত্ত এবং সংশ্লিষ্ট লক্ষণটি চল হলে সংগৃহীত উপাত্তকে বলা হবে পরিমাণগত উপাত্ত।

চল আবার দুধরনের হতে পারে। কিছু কিছু চল তাদের প্রসারসীমার মধ্যে মাত্র কয়েকটি বিচ্ছিন্ন মান গ্রহণ করতে পারে, কিন্তু প্রতিটি বা যে কোন মান গ্রহণ করে না। এই ধরনের চলকে বলা হয় **বিচ্ছিন্ন চল** (discrete variable)। যেমন, পরিবারের সদস্যসংখ্যা, বাড়ি-পিছু ঘরের সংখ্যা, ইত্যাদি। পরিবারের সদস্যসংখ্যা—এই চলটির মান কেবলমাত্র পূর্ণসংখ্যাই হতে পারে, $3\frac{1}{2}$ জন বা $15\frac{1}{4}$ জন সদস্য-বিশিষ্ট পরিবারের কথা আমরা ভাবতে পারি না।

আবার চলের সংজ্ঞার প্রসঙ্গে প্রদত্ত প্রথম পাঁচটি উদাহরণের চলগুলি তাদের সম্ভাব্য প্রসারসীমার মধ্যবর্তী যে কোন মান গ্রহণ করতে সক্ষম। যেমন, আলপিনের দৈর্ঘ্য 10 মিলিমিটার ( মি.মি. ) হতে পারে, 10.5 মি.মি. হওয়াও খুবই স্বাভাবিক এবং মাপন-যন্ত্রটি খুবই সূক্ষ্ম হলে বিশেষ কোন আলপিনের দৈর্ঘ্য 10.5364 মি.মি. পাওয়াও অসম্ভব কিছু নয়। এই ধরনের চলকে বলা হয় **সন্তত** বা **অবিচ্ছিন্ন** চল (continuous variable)। অবশ্য প্রচলিত মাপন-যন্ত্রের সীমাবদ্ধতার দরুণ কোনও অবিচ্ছিন্ন চলের ওপর সংগৃহীত মানগুলির মধ্যে কিছুটা বিচ্ছিন্নতা পরিলক্ষিত হওয়া খুবই সম্ভব। যেমন, ওজন ও উচ্চতা অবিচ্ছিন্ন চল—কিন্তু বাস্তবক্ষেত্রে মানুষের ওজন আসন্ন কিলোগ্রামে, সোনার ওজন আসন্ন মিলিগ্রামে অথবা মানুষের উচ্চতা আসন্ন সেন্টিমিটারেই পাওয়া যায়, এর থেকে সূক্ষ্মতর মাপ সাধারণ মাপন-যন্ত্রে পাওয়া যায় না বা পাওয়ার প্রয়োজন হয় না। আসলে বিচ্ছিন্ন চল এবং অবিচ্ছিন্ন চলের মধ্যে প্রকৃতিগত অন্তর্নিহিত তফাৎটি হ'ল, প্রথমোক্তটি একটি নির্দিষ্ট সীমার মধ্যে যে কোন মান গ্রহণ করতে পারে না, কিন্তু দ্বিতীয়টি পারে যদিও বাস্তবক্ষেত্রে কোনও অবিচ্ছিন্ন চল কর্তৃক গৃহীত মানগুলির মধ্যে কিছুটা কৃত্রিম বিচ্ছিন্নতা পরিলক্ষিত হওয়া সম্ভব।

## 3.3 পরিসংখ্যা-বিভাজন :

পরিসংখ্যা কথাটি 2.3 অনুচ্ছেদে কিছুটা আলোচিত হয়েছে। একাধিক ব্যষ্টির বিশেষ একটি লক্ষণের উপর উপাত্ত সংগ্রহ করার পর সংগৃহীত অবিন্যস্ত উপাত্ত পরিসংখ্যা-বিভাজন পদ্ধতির সাহায্যে সংক্ষেপিত করা যেতে পারে, একথা আগেই বলা হয়েছে। গুণ-লক্ষণ, বিচ্ছিন্ন চল এবং অবিচ্ছিন্ন চলের ক্ষেত্রে এই পদ্ধতির ওপর বিস্তারিত আলোচনা পরবর্তী কয়েকটি অনুচ্ছেদে করা হ'ল।

### 3.3.1 গুণ-লক্ষণের পরিসংখ্যা-বিভাজন :।

1971 সালের আদমশুমারির সময় পশ্চিমবঙ্গের কোনও একটি গ্রামের অধিবাসীদের সম্বন্ধে যে সব তথ্য সংগ্রহ করা হয়েছিল, তার মধ্যে একটি ছিল গ্রামবাসীদের শিক্ষাগত মান। এই উদ্দেশ্যে প্রাপ্তবয়স্ক গ্রামবাসীদের মোট চারটি শ্রেণীতে ভাগ করা হয়েছিল : নিরক্ষর, প্রাথমিক বিদ্যালয় মানের, মাধ্যমিক বিদ্যালয় মানের এবং মহাবিদ্যালয় ও বিশ্ববিদ্যালয় মানের। মোট 568 জন গ্রামবাসী সম্পর্কে তথ্যসংগ্রহ শেষ হলে ঐ গ্রামের জনৈক আগ্রহী

যুবক গ্রামের শিক্ষাগত মান সম্পর্কিত একটি পরিষ্কার চিত্র পাওয়ার উদ্দেশ্যে নিরক্ষর, প্রাথমিক বিদ্যালয় মানের,...ইত্যাদি চারটি শ্রেণীতে আগত গ্রামবাসীদের সংখ্যা গণনা ক'রে নিম্নলিখিত সারণীতে লিপিবদ্ধ করল :

## সারণী 3.1

## শিক্ষাগত মান অনুযায়ী 568 জন প্রাপ্তবয়স্ক গ্রামবাসীর পরিসংখ্যা-বিভাজন

| শিক্ষাগত মান | প্রাপ্তবয়স্ক গ্রামবাসীর সংখ্যা |
|---|---|
| নিরক্ষর | 321 |
| প্রাথমিক বিদ্যালয় মান | 153 |
| মাধ্যমিক বিদ্যালয় মান | 78 |
| কলেজ ও বিশ্ববিদ্যালয় মান | 16 |
| মোট | 568 |

এখানে যে গুণ-লক্ষণটি সম্পর্কে তথ্য আহরণ করা হয়েছে, সেটি হ'ল শিক্ষাগত মান। গুণ-লক্ষণটির মোট চারটি রূপ বা শ্রেণী নেওয়া হয়েছে এখানে। 568 জন প্রাপ্তবয়স্ক গ্রামবাসী সম্পর্কে এই গুণ-লক্ষণটির ওপর সংগৃহীত তথ্য সংক্ষেপীকরণের পর কেমন স্বল্পপরিসরে এবং সুশৃঙ্খলভাবে ওপরের সারণীতে লিপিবদ্ধ হয়েছে, লক্ষ্য কর। এই ধরনের সংক্ষেপীকরণের সময় গণনার সুবিধার জন্য **মিল-চিহ্নের** (tally-marks) ব্যবহার করা যেতে পারে। এ সম্পর্কে 3.3.2 অনুচ্ছেদে বিস্তারিত আলোচনা হবে। ওপরের সারণীতে 321 সংখ্যাটি সূচিত করছে 568 জনের মধ্যে মোট কতজন নিরক্ষর শ্রেণীভুক্ত—রাশিবিজ্ঞানের পরিভাষায় আমরা বলি 'শিক্ষাগত মান' এই গুণ-লক্ষণের 'নিরক্ষর' এই রূপটির পরিসংখ্যা হচ্ছে 321. বর্তমান ক্ষেত্রে 'মোট পরিসংখ্যা' 568. মোট পরিসংখ্যা যে ভাবে বিভিন্ন শ্রেণীর মধ্যে বিভাজিত বা নিবেশিত থাকে, তাকে বলা হয় **পরিসংখ্যা-বিভাজন** বা **পরিসংখ্যা-নিবেশন** (frequency distribution)। 3.1 সারণীতে এই বিভাজন বা নিবেশন প্রদর্শিত হয়েছে, তাই এটিকে বলা হয় **পরিসংখ্যা-বিভাজন** (বা **নিবেশন**) **সারণী** (frequency distribution table)।

এই প্রসঙ্গে লক্ষণীয়, যদিও গুণ-লক্ষণ সংখ্যার সাহায্যে প্রকাশ করা যায় না, তবুও লক্ষণভুক্ত বিভিন্ন শ্রেণীর পরিসংখ্যার হিসেব করতে গিয়ে আমাদের শেষ পর্যন্ত সংখ্যার আশ্রয় নিতেই হয়। অবশ্য চলের ক্ষেত্রে তথ্য সংগ্রহ সংখ্যা দিয়ে শুরু হয়, আর গুণ-লক্ষণের ক্ষেত্রে সংখ্যা আসে একেবারে শেষ পর্যায়ে।

গুণ-লক্ষণের বিভিন্ন রূপের আপেক্ষিক গুরুত্ব সম্বন্ধে স্পষ্টতর চিত্র পাওয়া যেতে পারে **আপেক্ষিক পরিসংখ্যা** (relative frequency) সারণী থেকে। কোন একটি শ্রেণীর আপেক্ষিক পরিসংখ্যা পাওয়া যায় ঐ শ্রেণীর পরিসংখ্যাকে [ এটিকে এখন **অনাপেক্ষিক পরিসংখ্যা** (absolute frequency) বলা যেতে পারে ] মোট পরিসংখ্যা দিয়ে ভাগ ক'রে। 3.1 সারণী থেকে নীচের আপেক্ষিক পরিসংখ্যা সারণীটি পাওয়া যেতে পারে :

**সারণী 3.2**

**468 জন গ্রামবাসীর মধ্যে বিভিন্ন মানের শিক্ষিতের অনুপাত**

| শিক্ষাগত মান | অনুপাত বা আপেক্ষিক পরিসংখ্যা |
|---|---|
| নিরক্ষর | ·5651 |
| প্রাথমিক বিদ্যালয় মান | ·2694 |
| মাধ্যমিক বিদ্যালয় মান | ·1373 |
| কলেজ ও বিশ্ববিদ্যালয় মান | ·0282 |
| মোট | 1·0000 |

একাধিক সমজাতীয় পরিসংখ্যা-বিভাজনের মধ্যে তুলনা করার সময়ও আপেক্ষিক পরিসংখ্যার সাহায্য নেওয়া হয়।

### 3.3.2 বিচ্ছিন্ন চলের পরিসংখ্যা-বিভাজন :

একটি উদাহরণ নিয়ে শুরু করা যাক। ইছাপুর হাই স্কুলে ( হুগলি ) 1968 সালে 20 জন শিক্ষক চাকুরীতে ছিলেন। ঐ বছরে স্কুল মোট 226 দিন খোলা ছিল। 3.3 সারণীতে স্কুল খোলা থাকার বিভিন্ন দিনে অনুপস্থিত শিক্ষকের সংখ্যা দেওয়া হয়েছে।

**সারণী 3.3**

**1968 সালের 226টি কার্যকালীন দিনে ইছাপুর হাই স্কুলে অনুপস্থিত শিক্ষকের সংখ্যা**

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 3 | 3 | 2 | 3 | 1 | 1 | 5 | 2 | 3 | 1 | 2 |
| 1 | 1 | 3 | 2 | 3 | 1 | 2 | 2 | 3 | 5 | 1 | 1 |
| 1 | 2 | 3 | 1 | 2 | 2 | 2 | 1 | 1 | 3 | 4 | 2 |
| 0 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 4 | 3 | 4 | 1 |
| 1 | 2 | 5 | 2 | 4 | 1 | 1 | 2 | 0 | 0 | 3 | 1 |
| 4 | 1 | 0 | 4 | 1 | 0 | 2 | 1 | 2 | 1 | 2 | 1 |
| 0 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 4 | 0 | 1 | 1 |
| 2 | 2 | 3 | 3 | 2 | 2 | 3 | 0 | 1 | 0 | 2 | 2 |
| 1 | 2 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 0 | 1 |
| 2 | 1 | 2 | 2 | 2 | 2 | 1 | 4 | 0 | 3 | 0 | 1 |
| 3 | 0 | 1 | 2 | 2 | 2 | 1 | 4 | 2 | 5 | 1 | 6 |
| 0 | 2 | 1 | 2 | 0 | 1 | 2 | 1 | 1 | 1 | 0 | 1 |
| 1 | 3 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 |
| 1 | 3 | 3 | 1 | 1 | 4 | 3 | 3 | 2 | 2 | 1 | 4 |
| 4 | 3 | 5 | 5 | 4 | 2 | 5 | 3 | 0 | 0 | 1 | 2 |
| 2 | 5 | 1 | 0 | 0 | 2 | 1 | 1 | 1 | 2 | 2 | 3 |
| 5 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |
| 1 | 2 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 |
| 1 | 0 | 1 | 1 | 0 | 0 | 1 | 2 | 0 | 1 | | |

উৎস : ইছাপুর হাই স্কুলের শিক্ষকদের হাজিরা বহি (1968)।

3.3 সারণীতে লিপিবদ্ধ বিপুল সংখ্যক অবিন্যস্ত রাশিতথ্যের প্রকৃতি এবং নানান বৈশিষ্ট্য সম্পর্কে সঠিক ধারণা পেতে হলে প্রথমেই দরকার সংশ্লিষ্ট চলটির (এখানে দিনপ্রতি অনুপস্থিত শিক্ষকদের সংখ্যা) পরিসংখ্যা-বিভাজন নিরূপণ করা। গুণ-লক্ষণের পরিসংখ্যা-বিভাজন নিরূপণের মতো এখানেও চলটির বিভিন্ন সম্ভাব্য মান কতবার ক'রে গৃহীত হয়েছে (অর্থাৎ, অনুপস্থিত শিক্ষকের সংখ্যা কতদিন 0, কতদিন 1,…ইত্যাদি), অর্থাৎ বিভিন্ন মানের পরিসংখ্যা গণনা ক'রে বের করা যেতে পারে। তবে আরও অল্প আয়াসে এবং সুশৃঙ্খলভাবে পরিসংখ্যা-বিভাজন নিরূপণ করার জন্য নিম্নলিখিত পদ্ধতিটি অনুসরণ করা যেতে পারে। প্রথমে দেখে নিতে হবে সংশ্লিষ্ট বিচ্ছিন্ন চলটি —ধরা যাক এটিকে আমরা $x$ দ্বারা চিহ্নিত করলাম—কী কী মান গ্রহণ করেছে।

আমাদের আলোচ্য উদাহরণে $x$ যে সব মান গ্রহণ করেছে সেগুলি হ'ল 0, 1, 2, 3, 4, 5 এবং 6। প্রস্তাবিত পরিসংখ্যা-বিভাজন সারণীর প্রথম স্তম্ভে লিপিবদ্ধ করতে হবে এই বিভিন্ন মানগুলি প্রত্যেক সারিতে একটি হিসাবে। এরপর অবিন্যস্ত রাশিতথ্যগুলির প্রথমটি থেকে পড়া শুরু করতে হবে এবং প্রতিটি মানের জন্য পরিসংখ্যা-বিভাজন সারণীর সংশ্লিষ্ট ঘরে একটি ক'রে মিলচিহ্ন দিতে হবে। গণনার সুবিধার জন্য একই ঘরে প্রতি পঞ্চম মিলচিহ্নটি পাশাপাশি না দিয়ে আগের চারটির ওপর কোনাকুনিভাবে বসিয়ে এক একটি পাঁচের স্তবক করা যেতে পারে ( সারণী 3.4 দ্রষ্টব্য )। এইভাবে অবিন্যস্ত রাশিতথ্যের সবটুকু মিলচিহ্নে রূপান্তরিত হয়ে গেলে খুব সহজেই পরিসংখ্যাগুলি বের করা যায় এবং এগুলি দেখানো হয় পরবর্তী স্তম্ভে।

**সারণী 3.4**

**দিনপ্রতি অনুপস্থিত শিক্ষকদের পরিসংখ্যা-বিভাজন নির্ণয়**

| অনুপস্থিত শিক্ষকের সংখ্যা | মিলচিহ্ন | দিনের সংখ্যা (পরিসংখ্যা) |
|---|---|---|
| 0 | 卌 卌 卌 卌 卌 卌 卌 卌 | 40 |
| 1 | 卌 卌 卌 卌 卌 卌 卌 卌 卌 | |
| | 卌 卌 卌 卌 卌 卌 卌 II | 82 |
| 2 | 卌 卌 卌 卌 卌 卌 卌 卌 卌 | |
| | 卌 卌 II | 57 |
| 3 | 卌 卌 卌 卌 IIII | 24 |
| 4 | 卌 卌 III | 13 |
| 5 | 卌 IIII | 9 |
| 6 | I | 1 |
| মোট | — | 226 |

ওপরের সারণী থেকে $x$-এর বিভিন্ন মানের পরিসংখ্যাগুলি সরাসরি পাওয়া যায়। আগের নিয়মেই বিভিন্ন মানের আপেক্ষিক পরিসংখ্যাও পাওয়া যেতে পারে।

অনেক সময় আবার দিনপ্রতি 3 জন অথবা তার কম, কিংবা 2 জন অথবা তার বেশী সংখ্যক শিক্ষক অনুপস্থিত হয়েছেন কতদিন, ইত্যাদি জানা প্রয়োজন হয়। সরাসরি এই জাতীয় প্রশ্নের উত্তর পেতে হলে **ক্রমযৌগিক পরিসংখ্যা** (cumulative frequency) নির্ণয় করা দরকার। $x$-এর বিভিন্ন মানের পরিসংখ্যাগুলি পর পর যোগ ক'রে পাওয়া যায় ক্রমযৌগিক পরিসংখ্যা। চলটির সর্বনিম্ন মান থেকে শুরু ক'রে অন্যান্য মান পর্যন্ত পরিসংখ্যাগুলির ক্রমিক যোগফল হচ্ছে চলটির এক-একটি 'ক্ষুদ্রতর-সূচক ক্রমযৌগিক পরিসংখ্যা' (cumulative frequency of less-than type)। পক্ষান্তরে চলটির সর্বোচ্চ মান থেকে শুরু ক'রে অন্যান্য মান পর্যন্ত পরিসংখ্যাগুলির ক্রমিক যোগফলকে বলা হয় চলটির এক-একটি 'বৃহত্তর-সূচক ক্রমযৌগিক পরিসংখ্যা' (cumulative frequency of more-than type)। অনুরূপভাবে, ক্ষুদ্রতর-সূচক এবং বৃহত্তর-সূচক ক্রমযৌগিক আপেক্ষিক পরিসংখ্যা বা অনুপাত বা শতকরা হারও পাওয়া যেতে পারে। 3.3 সারণীতে প্রদত্ত রাশিতথ্যের পরিসংখ্যা-বিভাজন, ক্রমযৌগিক পরিসংখ্যা-বিভাজন এবং ক্রমযৌগিক আপেক্ষিক পরিসংখ্যা-বিভাজন দেখানো হয়েছে 3.5 সারণীতে। এই সারণী থেকে সরাসরি পাওয়া যায়, দিনপ্রতি 3 জন অথবা তার কম সংখ্যক শিক্ষক অনুপস্থিত হয়েছেন 203 দিন, কিংবা দিনপ্রতি 2 বা তার বেশী সংখ্যক শিক্ষক অনুপস্থিত হয়েছেন 104 দিন, ইত্যাদি। আরও পাওয়া যায়, দিনপ্রতি 3 অথবা তার কম এবং দিনপ্রতি 2 অথবা তার বেশী সংখ্যক শিক্ষক অনুপস্থিত হয়েছেন যথাক্রমে শতকরা 89·82 দিন এবং শতকরা 46·02 দিন, ইত্যাদি।

## সারণী 3.5

### দিনপ্রতি অনুপস্থিত শিক্ষকদের পরিসংখ্যা-বিভাজন, আপেক্ষিক পরিসংখ্যা-বিভাজন এবং ক্রমযৌগিক পরিসংখ্যা-বিভাজন

| অনুপস্থিত শিক্ষকের সংখ্যা | পরিসংখ্যা | আপেক্ষিক পরিসংখ্যা ( শতকরা ) | ক্রমযৌগিক পরিসংখ্যা | | ক্রমযৌগিক শতকরা হার | |
|---|---|---|---|---|---|---|
| | | | ক্ষুদ্রতর-সূচক | বৃহত্তর-সূচক | ক্ষুদ্রতর-সূচক | বৃহত্তর-সূচক |
| 0 | 40 | 17·70 | 40 | 226 | 17·70 | 100·00 |
| 1 | 82 | 36·28 | 122 | 186 | 53·98 | 82·30 |
| 2 | 57 | 25·22 | 179 | 104 | 79·20 | 46·02 |
| 3 | 24 | 10·62 | 203 | 47 | 89·82 | 20·80 |
| 4 | 13 | 5·76 | 216 | 23 | 95·58 | 10·18 |
| 5 | 9 | 3·98 | 225 | 10 | 99·56 | 4·42 |
| 6 | 1 | 0·44 | 226 | 1 | 100·00 | 0·44 |
| মোট | 226 | 100·00 | – | – | – | – |

এখানে $x$-এর সম্ভাব্য মানের মোট সংখ্যা খুব বেশী না হওয়ায় গৃহীত প্রতিটি মানের জন্য পৃথক্ পৃথক্ ভাবে পরিসংখ্যা বের করা সম্ভব হয়েছে। সম্ভাব্য মানের সংখ্যা বেশী হলে একাধিক মান-সম্বলিত এক-একটি শ্রেণী গঠন ক'রে বিভিন্ন শ্রেণীর পরিসংখ্যা নির্ণয় করতে হবে।

### 3.3.3 অবিচ্ছিন্ন চলের পরিসংখ্যা-বিভাজন :

গুণলক্ষণ এবং বিচ্ছিন্ন চলের ক্ষেত্রে পরিসংখ্যা-বিভাজন নিরূপণের উদ্দেশ্যে শ্রেণীবিন্যাসের প্রশ্নটি কিছুটা সরলীকৃত, কেননা এখানে বিভিন্ন শ্রেণীতে প্রায়ই চলের এক একটি মান বা গুণলক্ষণের এক একটি রূপ নেওয়া হয়ে থাকে, অথবা বলা যায় সংগৃহীত রাশিতথ্যের প্রকৃতি থেকেই শ্রেণীবিন্যাসের ধারা সম্বন্ধে একটি স্বচ্ছ ধারণা পাওয়া যায়। কিন্তু অবিচ্ছিন্ন চলের ক্ষেত্রে সমস্যাটি কিছুটা জটিল, কারণ সেক্ষেত্রে সম্ভাব্য মানের সংখ্যা অসীম হওয়ায় এক-একটি মানের জন্য এক-একটি শ্রেণী গ্রহণ করা যায় না। সুতরাং এক্ষেত্রে প্রয়োজন, চলের প্রসারটি কয়েকটি-

ভাগে ভাগ ক'রে এক-একটি ভাগকে এক-একটি শ্রেণীরূপে গণ্য করা। স্পষ্টতঃই এক্ষেত্রে শ্রেণীবিন্যাস কিছুটা কৃত্রিম হতে বাধ্য এবং বিশেষ ক্ষেত্রে কী ধরনের শ্রেণীবিন্যাস যথাযথ হবে তা নির্ভর করে সংগৃহীত রাশিতথ্যের সংখ্যা, প্রসার, প্রকৃতি ইত্যাদির ওপর এবং সবার ওপর শ্রেণীবিন্যাসকারীর দক্ষতা এবং বিচার-বুদ্ধির ওপর।

একটি উদাহরণ নিয়ে শুরু করা যাক। 3.6 সারণীতে কলকাতায় 1972 সালের মার্চ মাস থেকে জুলাই মাস পর্যন্ত এই 5 মাসের দৈনন্দিন সর্বোচ্চ তাপমাত্রা ফারেনহাইট ডিগ্রিতে দেওয়া হয়েছে। সাধারণতঃ গ্রীষ্মকালে কলকাতার দৈনন্দিন তাপমাত্রার কী ধরনের তারতম্য ঘটে সে সম্বন্ধে একটি পরিষ্কার চিত্র পাওয়া যাবে এ থেকে। গ্রীষ্মে এবং শীতে এই চিত্র লক্ষণীয়ভাবে ভিন্ন। তাই মোটামুটি অন্তঃসম (homogeneous) তথ্যের উদাহরণ হিসাবে শুধুমাত্র গ্রীষ্মকালের তাপমাত্রাই নেওয়া হয়েছে।

## সারণী 3.6

### কলকাতার দৈনন্দিন সর্বোচ্চ তাপমাত্রা ( ডিঃ ফাঃ )

### [ মার্চ—জুলাই, 1972 ]

| | | | | | | |
|---|---|---|---|---|---|---|
| মার্চ। | 86·0 | 85·8 | 86·7 | 88·7 | 88·9 | 89·4 |
| | 88·9 | 85·4 | 90·3 | 91·4 | 93·4 | 92·7 |
| | 92·4 | 93·7 | 92·5 | 90·8 | 90·5 | 95·7 |
| | 94·8 | 94·1 | 97·9 | 98·4 | 97·7 | 97·7 |
| | 99·6 | 95·5 | 95·5 | 99·1 | 91·9 | 95·9 |
| | 96·3 | | | | | |
| এপ্রিল। | 98·1 | 93·5 | 97·5 | 95·7 | 94·3 | 94·9 |
| | 95·9 | 97·5 | 93·4 | 95·7 | 91·8 | 96·1 |
| | 95·3 | 97·7 | 97·1 | 97·8 | 94·8 | 92·5 |
| | 103·8 | 95·7 | 94·9 | 100·8 | 103·7 | 104·7 |
| | 104·5 | 105·6 | 102·0 | 95·5 | 97·1 | 97·8 |
| মে। | 105·6 | 100·9 | 100·2 | 100·4 | 103·6 | 101·6 |
| | 107·6 | 103·0 | 106·6 | 102·1 | 103·6 | 104·2 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | 104·7 | 101·1 | 99·1 | 95·7 | 97·7 | 99·5 |
| | 101·2 | 106·7 | 98·0 | 101·3 | 99·9 | 93·2 |
| | 91·9 | 91·9 | 92·8 | 98·4 | 97·7 | 98·5 |
| | 97·7 | | | | | |
| জুন। | 95·9 | 96·8 | 95·7 | 97·4 | 98·6 | 102·0 |
| | 105·8 | 103·8 | 107·1 | 101·7 | 104·0 | 102·4 |
| | 98·4 | 93·2 | 95·5 | 95·4 | 94·7 | 94·9 |
| | 91·4 | 94·2 | 94·0 | 95·9 | 88·5 | 89·4 |
| | 88·5 | 88·7 | 90·7 | 92·3 | 93·0 | 95·0 |
| জুলাই। | 89·3 | 88·7 | 92·3 | 91·9 | 93·1 | 92·1 |
| | 91·4 | 87·6 | 86·7 | 90·5 | 96·6 | 96·6 |
| | 101·9 | 92·3 | 90·0 | 85·9 | 90·4 | 88·7 |
| | 83·2 | 88·5 | 93·6 | 94·6 | 91·9 | 91·2 |
| | 92·8 | 92·8 | 95·7 | 94·1 | 94·1 | 96·2 |
| | 92·3 | | | | | |

উৎস : দৈনিক স্টেট্‌স্‌ম্যান : মার্চ—জুলাই, 1972.

অবিচ্ছিন্ন চলের শ্রেণীবিন্যাস সম্বন্ধে যদিও কোন ধরাবাঁধা নিয়ম নেই, তবুও সাধারণভাবে কয়েকটি কথা বলা যেতে পারে।

প্রথমতঃ, গৃহীত শ্রেণীগুলি অর্থাৎ চলটির মোট প্রসারকে যে কয়েকটি উপ-প্রসারে ভাগ করা হবে, সেগুলি পরস্পর-নিঃশেষী (mutually exhaustive) হওয়া প্রয়োজন—অর্থাৎ সারণীগত প্রতিটি মান যেন কোন না কোন শ্রেণীর অন্তর্গত হয়। এর জন্য প্রথমতঃ খুঁজে বের করতে হবে সারণীতে প্রদত্ত ক্ষুদ্রতম এবং বৃহত্তম মান-দুটি এবং দেখতে হবে, গৃহীত প্রথম শ্রেণীটির অধঃসীমা যেন এই ক্ষুদ্রতম মানটি থেকে বেশী না হয় এবং সর্বশেষ শ্রেণীটির উর্ধ্বসীমা এই বৃহত্তম মানটি থেকে কম না হয়। আমাদের বর্তমান উদাহরণে ক্ষুদ্রতম মান 83·2 এবং বৃহত্তম মান 107·6—সুতরাং নিম্নতম শ্রেণী 82·0 থেকে শুরু করা যেতে পারে এবং উর্ধ্বতম শ্রেণী শেষ করা যেতে পারে 107·9-এ।

দ্বিতীয়তঃ, শ্রেণীগুলিকে পরস্পর-বিচ্ছিন্ন (mutually exclusive) হতে হবে, অর্থাৎ কোন দুটি শ্রেণী নেওয়া হলে চলটির প্রসারের সামান্যতম অংশও একই সঙ্গে

যুগপৎ এই দুটি শ্রেণীরই অন্তর্ভুক্ত হওয়া চলবে না। যেমন 32·0—85·0, 85·0—88·0, ......এইভাবে শ্রেণীগুলি নির্বাচন করা হলে সারণীগত 85·0 মানটি প্রথম শ্রেণীতে না দ্বিতীয় শ্রেণীতে অন্তর্ভুক্ত করা হবে তাই নিয়ে অনিশ্চয়তা দেখা দেবে। এই ধরনের অসুবিধা এড়ানোর জন্য শ্রেণীগুলিকে 82·0—, 85·0—, 88·0—, ......এইভাবে নির্বাচন করা যেতে পারে। এর অর্থ 82·0 অথবা তার বেশী কিন্তু 85·0 এর কম যে কোন মান অন্তর্ভুক্ত হবে প্রথম শ্রেণীটিতে, 85·0 অথবা তার বেশী কিন্তু 88·0 এর কম মানগুলি অন্তর্ভুক্ত হবে দ্বিতীয় শ্রেণীতে, ইত্যাদি। অথবা, প্রদত্ত মানগুলি কত দশমিক স্থান পর্যন্ত দেওয়া আছে সেদিকে খেয়াল রেখে শ্রেণীগুলি শুরুতেই পরস্পর-বিচ্ছিন্ন ক'রে নেওয়া যেতে পারে। যেমন প্রদত্ত সবকটি মান অখণ্ড সংখ্যায় দেওয়া থাকলে—যেমন 100 নম্বর পূর্ণসংখ্যাযুক্ত কোন পরীক্ষায় ছাত্রছাত্রীদের প্রাপ্ত নম্বর ( সবসময়েই অখণ্ড সংখ্যায় প্রকাশিত ধ'রে নিয়ে )—শ্রেণীগুলি 0—9, 10—19, ......এইভাবে নেওয়া যেতে পারে। তেমনি বর্তমান উদাহরণে মানগুলি দেওয়া আছে আসন্ন প্রথম দশমিকে—এখানে 82·0—84·9, 85·0—87·9, ......এইভাবে শ্রেণীগুলি নির্বাচিত হতে পারে। অথবা বর্তমান উদাহরণে যদি মানগুলি আসন্ন দ্বিতীয় দশমিকে দেওয়া হ'ত তাহলে উপযুক্ত শ্রেণীবিন্যাস হ'ত 82·00—84·99, 85·00—87·99, ......এইভাবে।

তৃতীয়তঃ, শ্রেণীসংখ্যা খুব বেশী অথবা খুব কম হওয়া উচিত নয়। শ্রেণী-সংখ্যা খুব কম হওয়ার অর্থ হ'ল শ্রেণীদৈর্ঘ্যগুলিকে বাড়িয়ে দেওয়া। এর প্রথম অসুবিধা, এর ফলে মূল রাশিতথ্যের কিছু কিছু প্রয়োজনীয় বিচ্যুতি সংক্ষেপিত রাশিতথ্যে চোখে না পড়ার সম্ভাবনা। এ ছাড়া আরও একটি গুরুত্বপূর্ণ অসুবিধা আছে। শ্রেণীবিন্যাসযুক্ত সংক্ষেপিত রাশিতথ্য থেকে জানা যায় 85·0—87·9 এই শ্রেণীটিতে আছে, ধরা যাক, মোট 4 দিনের সর্বোচ্চ তাপমাত্রা—কিন্তু এথেকে এই 4 দিনের সর্বোচ্চ তাপমাত্রার সঠিক পরিমাণ আর জানা সম্ভব নয়। অথচ এই রাশিতথ্য প্রয়োজনানুগভাবে আরও বিশ্লেষণ করতে হলে শ্রেণী-অন্তর্ভুক্ত সঠিক মানগুলি সম্বন্ধে একটা ধারণা করতেই হয়। এই উদ্দেশ্যে কোন শ্রেণীর পরিসংখ্যা শ্রেণীটির ওপর সমভাবে বিন্যস্ত ধ'রে নিয়ে শ্রেণীর অন্তর্গত একটি বিশেষ মানকে ( যেমন মধ্যবিন্দুটি ) শ্রেণীটির প্রতিনিধিস্থানীয় মান হিসাবে গ্রহণ করা হয় এবং এই শ্রেণীতে আগত সবকটি মানই এই প্রতিভূ মানের সমান ধ'রে নেওয়া হয়। এখন শ্রেণী-অন্তরটি খুব বড় হলে স্পষ্টতঃই এই স্বীকরণ বাস্তব-

সম্মত না হওয়ার সম্ভাবনা বাড়ে। আবার শ্রেণীসংখ্যা খুব বেশী হলেও শ্রেণীবিন্যাস অযথা জটিল হয়ে ওঠে, শ্রেণীবিন্যাসের মূল উদ্দেশ্যটিই হয় ব্যাহত। তাছাড়া সেক্ষেত্রে শ্রেণীবিন্যস্ত রাশিতথ্যে এমন সব অনিয়মিত ধাঁচ (irregular pattern) দেখা দিতে পারে যেগুলি আসল পরিসংখ্যা-বিভাজনে হয়তো একেবারেই অনুপস্থিত। এ বিষয়ে সাধারণ নিয়ম হ'ল, মোট পরিসংখ্যা 1000 বা তার বেশী হলে মোট 10 থেকে 20টি শ্রেণী নেওয়া যেতে পারে। মোট পরিসংখ্যা 1000 থেকে অনেক কম হলে অবশ্য আরও কম সংখ্যক শ্রেণী নেওয়াই যুক্তিযুক্ত।

চতুর্থতঃ, শ্রেণীগুলি সমদৈর্ঘ্য হওয়া বাঞ্ছনীয়। এর ফলে পরবর্তী পর্যায়ে রাশিতথ্য বিশ্লেষণে বিশেষ সুবিধা হয়। অবশ্য বিশেষ বিশেষ ক্ষেত্রে অসমদৈর্ঘ্য শ্রেণীবিন্যাস অপরিহার্য হয়ে পড়ে। যেমন অনেক সময় এমন হতে পারে, চলটির প্রসার খুব বিরাট, অথচ গৃহীত মানগুলি প্রসারের বিশেষ একটি অঞ্চলে (প্রথম কিংবা শেষ দিকে, অথবা অন্য কোন বিশেষ অঞ্চলে) বেশী কেন্দ্রীভূত, অবশিষ্ট বেশীরভাগ অংশে খুব কম। এক্ষেত্রে শেষোক্ত অঞ্চলের শ্রেণীগুলির তুলনায় প্রথমোক্ত অঞ্চলের শ্রেণীগুলির দৈর্ঘ্য কম নেওয়াই যুক্তিযুক্ত। 3.8 সারণীতে একটি অসমদৈর্ঘ্য শ্রেণীবিন্যাসের উদাহরণ দেওয়া হয়েছে।

বর্তমান উদাহরণে মোট পরিসংখ্যা 153—এক্ষেত্রে 9টি শ্রেণী নেওয়া খুব অযৌক্তিক হবে না। মূল প্রসার 24·4 (=107·6—83·2), অর্থাৎ 27-এর কিছু কম, সুতরাং এক-একটি শ্রেণীর দৈর্ঘ্য 3 একক হিসাবে ধ'রে সমদৈর্ঘ্য বিভিন্ন শ্রেণীগুলি নেওয়া যেতে পারে 82·0—84·9, 85·0—87·9, ..., 106·0—108·9. শ্রেণীগুলি নির্ধারিত হওয়ার পর আগের মত মিলচিহ্নের সাহায্যে বিভিন্ন শ্রেণী-পরিসংখ্যাগুলি পাওয়া যেতে পারে (সারণী 3.7)।

## সারণী 3.7

### কলকাতার দৈনন্দিন সর্বোচ্চ তাপমাত্রার পরিসংখ্যা-বিভাজন নিরূপণ

### [ মার্চ—জুলাই, 1972 ]

| দৈনন্দিন সর্বোচ্চ তাপমাত্রা [ ডিঃ ফাঃ ] | মিলচিহ্ন | পরিসংখ্যা |
|---|---|---|
| 82·0— 84·9 | \| | 1 |
| 85·0— 87·9 | 𝍸 \|\| | 7 |
| 88·0— 90·9 | 𝍸 𝍸 𝍸 \|\|\|\| | 19 |
| 91·0— 93·9 | 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 \| | 31 |
| 94·0— 96·9 | 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 \|\| | 37 |
| 97·0— 99·9 | 𝍸 𝍸 𝍸 𝍸 𝍸 \| | 26 |
| 100·0—102·9 | 𝍸 𝍸 \|\|\|\| | 14 |
| 103·0—105·9 | 𝍸 𝍸 \|\|\|\| | 14 |
| 106·0—108·9 | \|\|\|\| | 4 |
| মোট | — | 153 |

ওপরের আলোচনা থেকে স্পষ্টই প্রতীয়মান হচ্ছে যে একই মূল অবিন্যস্ত রাশিতথ্য থেকে বিভিন্ন জনে বিভিন্ন ধরনের শ্রেণীবিন্যাস করতে পারে।

### সারণী 3.8

### লোকসংখ্যা অনুযায়ী পশ্চিমবঙ্গে গ্রামের পরিসংখ্যা-বিভাজন ( 1961 সালে )

| লোকসংখ্যা | গ্রামের সংখ্যা |
|---|---|
| 500-এর কম | 22,291 |
| 500— 999 | 8,514 |
| 1,000—1,999 | 5,224 |
| 2,000—4,999 | 2,156 |
| 5,000—9,999 | 244 |
| 10,000 এবং তদূর্ধ্ব | 25 |
| মোট | 38,454 |

উৎস : *Census of India, 1961 ; Vol. XVI, Part IIA.*

**কয়েকটি সংজ্ঞা :**

**শ্রেণী-অন্তর** (class-interval), **শ্রেণী-সীমা** (class-limits) ও **শ্রেণী-সীমান্ত** (class-boundaries) :

ওপরের উদাহরণে 82·0—, 85·0—, ..., অথবা 82·0—84·9, 85·0—87·9, ......এই মানসীমাগুলিকে এক একটি **শ্রেণী-অন্তর** বলে। এখানে তাপমাত্রা লিপিবদ্ধ করা হয়েছে আসন্ন দশমাংশে। সুতরাং 82·0—, এই শ্রেণীতে অন্তর্ভুক্ত করা হবে মাত্র 82·0, 82·1, ...84·9 এই ক'টি মান। অর্থাৎ প্রকৃতপক্ষে 82·0 থেকে 84·9-এর মধ্যেই থাকছে আলোচ্য শ্রেণী-অন্তরটির বিভিন্ন মান। এর অর্থ হ'ল, যদিও 82·0 বা তার বেশি কিন্তু 85·0-এর কম যে কোন মান এই শ্রেণীভুক্ত হওয়ার কথা, কার্যতঃ 84·9-এর বেশী কোন মান এই শ্রেণীর অন্তর্ভুক্ত করা হচ্ছে না। তাই 82·0 এবং 84·9-কে শ্রেণী-অন্তরটির যথাক্রমে আপাত অধঃসীমা এবং উর্ধ্বসীমা বা শুধু **অধঃসীমা** এবং **উর্ধ্বসীমা** বলা হয়।

এখন এখানে তাপমাত্রা আসন্ন দশমাংশে লিপিবদ্ধ করা হয়েছে ব'লে 81·95 থেকে 82·05 পর্যন্ত যে কোন মানকেই লেখা হয়েছে 82·0 হিসাবে। তেমনি 84·85 থেকে 84·95 পর্যন্ত যে কোন মানকে ধরা হয়েছে 84·9 হিসাবে। সুতরাং

82·0—84·9 এই শ্রেণী-অন্তরটির প্রকৃত শ্রেণী-সীমা হচ্ছে 81·95—84·95. 81·95 ও 84·95-কে যথাক্রমে শ্রেণী-অন্তরটির **অধঃ-** এবং **উর্ধ্ব শ্রেণী-সীমান্ত** বলা হয়।

স্পষ্টতঃই শ্রেণী-অন্তর শ্রেণী-সীমান্তেও দেওয়া হতে পারে, যেমন আমরা বলতে পারি 81·95—84·95 এই শ্রেণী-অন্তরটি। শ্রেণী-অন্তর সীমায় না সীমান্তে দেওয়া আছে, তা বুঝতে হলে দেখতে হবে কোন শ্রেণীর উর্ধ্বমান এবং পরবর্তী শ্রেণীর নিম্নমানের মধ্যে কোন ফাঁক আছে কি না। ফাঁক থাকলে বুঝতে হবে শ্রেণীটি দেওয়া আছে সীমায়, অন্যথায় সীমান্তে। কোন শ্রেণীর সীমা থেকে সীমান্ত পেতে হলে এই ফাঁকটুকুর পরিমাণ বের করতে হবে, তারপর এই পরিমাণের অর্ধাংশ বাদ দিতে হবে অধঃসীমা থেকে, বাকী অর্ধাংশ যোগ করতে হবে উর্ধ্বসীমার সঙ্গে। যেমন 82·0—84·9, 85·0—87·9, ... এই শ্রেণী-অন্তরগুলি দেওয়া আছে সীমায়। এখানে দুটি পাশাপাশি অন্তরের মধ্যে ফাঁকের পরিমাণ 0·1—সুতরাং শ্রেণী-সীমা থেকে পাওয়া সীমান্তগুলি হবে 81·95(=82·0−·05)—84·95(=84·9+·05), 84·95—87·95, ... ইত্যাদি। তেমনি 10—19, 20—29, ... ইত্যাদি শ্রেণী-সীমায় দেওয়া অন্তরগুলি শ্রেণী-সীমান্তে লেখা হলে দাঁড়াবে 9·5—19·5, 19·5—29·5,... ইত্যাদি। লক্ষণীয়, সীমাগুলির মধ্যবর্তী ফাঁকটুকু বস্তুতঃ কৃত্রিম—এটি মাপনযন্ত্রের সীমাবদ্ধতা (limitation) থেকে উদ্ভূত।

অবিন্যস্ত রাশিতথ্য সারণীবদ্ধ করার প্রয়োজনে শ্রেণীগুলি শ্রেণী-সীমায় নেওয়াই সুবিধাজনক। কিন্তু পরবর্তী বিশ্লেষণের সময় শ্রেণী-সীমান্তগুলিই বেশী প্রয়োজনীয় হয়ে দাঁড়ায়।

অনেক সময় রাশিতথ্য শ্রেণীবিন্যস্ত আকারেই দেওয়া থাকে সেক্ষেত্রে মূল মানগুলি যত দশমিক স্থান পর্যন্ত আছে, শ্রেণীগুলি তত দশমিক স্থান পর্যন্ত সীমান্তে প্রদত্ত হলে প্রশ্ন জাগতে পারে, সীমান্তবর্তী মানগুলি কোন্ শ্রেণীতে নেওয়া হয়েছে। এক্ষেত্রে বিভিন্ন রীতি প্রচলিত—কেউ কেউ এইসব মান পূর্ববর্তী শ্রেণীতে নেওয়ার পক্ষপাতী, কেউ আবার এগুলি পরবর্তী শ্রেণীতে নিয়ে থাকেন। তৃতীয় আর একটি প্রচলিত রীতি হ'ল উভয় শ্রেণীতেই অর্ধাংশ পরিমাণ পরিসংখ্যা নেওয়া। যাই হোক, শ্রেণীবিন্যাসের ক্ষেত্রে একই সংখ্যক দশমিক পর্যন্ত শ্রেণী-সীমান্ত নেওয়া বাঞ্ছনীয় নয়।

**শ্রেণী-মধ্যক** (class mid-point or class-mark) :

শ্রেণী-অন্তরের উর্ধ্বমান এবং অধঃমানের ( সীমা কিংবা সীমান্ত ) যোগফলের

অর্ধাংশ হ'ল শ্রেণী-মধ্যক। যেমন 82·0—84·9 এই শ্রেণীটির মধ্যক হ'ল $\frac{1}{2}(82.0+84.9)=83.45=\frac{1}{2}(81.95+84.85)$। শ্রেণী-মধ্যককে শ্রেণী-অন্তরটির প্রতিনিধিস্থানীয় মান বলা হয়। স্পষ্টতঃই সমদৈর্ঘ্য শ্রেণীবিন্যাসে শ্রেণী-মধ্যকগুলিও সমদূরস্থিত হয়।

**শ্রেণী-দৈর্ঘ্য** বা **শ্রেণী-প্রসার** (class width) : কোন শ্রেণীর উর্ধ্বসীমান্ত এবং অধঃসীমান্তের বিয়োগফল হ'ল ঐ শ্রেণী-অন্তরের দৈর্ঘ্য। লক্ষণীয়, পরবর্তী শ্রেণীর অধঃসীমা থেকে সংশ্লিষ্ট শ্রেণীর অধঃসীমা বিয়োগ ক'রেও শ্রেণীটির দৈর্ঘ্য পাওয়া যায়। 82·0 – 84·9 এই শ্রেণী-অন্তরটির দৈর্ঘ্য হবে $84.95-81.95=3$ $(=85.0-82.0)$ ডিঃ ফাঃ।

**পরিসংখ্যা-ঘনত্ব** (frequency density) : কোন শ্রেণী-অন্তরের পরিসংখ্যাকে দৈর্ঘ্য দিয়ে ভাগ ক'রে পাওয়া যায় ঐ শ্রেণী-অন্তরের পরিসংখ্যা-ঘনত্ব, বা প্রতি এককে পরিসংখ্যার পরিমাণ। অসমদৈর্ঘ্য শ্রেণী-বিন্যাসের ক্ষেত্রে বিভিন্ন শ্রেণীর পরিসংখ্যাগুলি পরস্পর তুলনীয় নয়, কিন্তু পরিসংখ্যা-ঘনত্বগুলি পরস্পর তুলনীয়।

3.9 সারণীতে আমাদের বর্তমান উদাহরণটির বিভিন্ন শ্রেণী-অন্তরের সীমা, সীমান্ত, মধ্যক, দৈর্ঘ্য, পরিসংখ্যা-ঘনত্ব, ইত্যাদি দেওয়া হয়েছে।

## সারণী 3.9

### কলকাতার দৈনন্দিন সর্বোচ্চ তাপমাত্রার পরিসংখ্যা-বিভাজন নিরূপণ [ মার্চ—জুলাই, 1972 ]

| দৈনন্দিন সর্বোচ্চ তাপমাত্রা ( ডিঃ ফাঃ ) | | শ্রেণী-মধ্যক ( ডিঃ ফাঃ ) | শ্রেণী-দৈর্ঘ্য ( ডিঃ ফাঃ ) | পরিসংখ্যা | পরিসংখ্যা-ঘনত্ব |
|---|---|---|---|---|---|
| শ্রেণী-সীমা | শ্রেণী-সীমান্ত | | | | |
| 82·0— 84·9 | 81·95— 84·95 | 83·45 | 3 | 1 | 0·33 |
| 85·0— 87·9 | 84·95— 87·95 | 86·45 | 3 | 7 | 2·33 |
| 88·0— 90·9 | 87·95— 90·95 | 89·45 | 3 | 19 | 6·33 |
| 91·0— 93·9 | 90·95— 93·95 | 92·45 | 3 | 31 | 10·33 |
| 94·0— 96·9 | 93·95— 96·95 | 95·45 | 3 | 37 | 12·33 |
| 97·0— 99·9 | 96·95— 99·95 | 98·45 | 3 | 26 | 8·67 |
| 100·0—102·9 | 99·95—102·95 | 101·45 | 3 | 14 | 4·67 |
| 103·0—105·9 | 102·95—105·95 | 104·45 | 3 | 14 | 4·67 |
| 106·0—108·9 | 105·95—108·95 | 107·45 | 3 | 4 | 1·33 |
| মোট | — | — | — | 153 | — |

অবিচ্ছিন্ন চলের ক্ষেত্রেও ক্রমযৌগিক পরিসংখ্যা পাওয়া যেতে পারে। প্রথম শ্রেণী থেকে শুরু ক'রে অন্য কোন শ্রেণী পর্যন্ত পরিসংখ্যার যোগফল হ'ল শ্রেণীটির ক্ষুদ্রতর-সূচক ক্রমযৌগিক পরিসংখ্যা। তেমনি কোন একটি শ্রেণী থেকে শুরু ক'রে সর্বশেষ শ্রেণীটি পর্যন্ত পরিসংখ্যাগুলির যোগফল শ্রেণীটির বৃহত্তর-সূচক ক্রমযৌগিক পরিসংখ্যা। 3.10 সারণীতে আলোচ্য উদাহরণের ক্রমযৌগিক পরিসংখ্যা এবং শতকরা হারের বিভাজন দেখানো হয়েছে।

**সারণী 3.10**

**কলকাতার দৈনন্দিন সর্বোচ্চ তাপমাত্রার পরিসংখ্যা-বিভাজন**

**[ মার্চ—জুলাই, 1972 ]**

| দৈনন্দিন তাপমাত্রা (ডিঃ ফাঃ) | পরিসংখ্যা | পরিসংখ্যার শতকরা হার | ক্রমযৌগিক পরিসংখ্যা | | ক্রমযৌগিক শতকরা হার | |
|---|---|---|---|---|---|---|
| | | | ক্ষুদ্রতর-সূচক | বৃহত্তর-সূচক | ক্ষুদ্রতর-সূচক | বৃহত্তর-সূচক |
| 82·0— 84·9 | 1 | 0·65 | 1 | 153 | 0·65 | 100·00 |
| 85·0— 87·9 | 7 | 4·58 | 8 | 152 | 5·23 | 99·35 |
| 88·0— 90·9 | 19 | 12·42 | 27 | 145 | 17·65 | 94·77 |
| 91·0— 93·9 | 31 | 20·26 | 58 | 126 | 37·91 | 82·35 |
| 94·0— 96·9 | 37 | 24·18 | 95 | 95 | 62·09 | 62·09 |
| 97·0— 99·9 | 26 | 16·99 | 121 | 58 | 79·08 | 37·91 |
| 100·0—102·9 | 14 | 9·15 | 135 | 32 | 88·23 | 20·92 |
| 103·0—105·9 | 14 | 9·16 | 149 | 18 | 97·38 | 11·77 |
| 106·0—108·9 | 4 | 2·62 | 153 | 4 | 100·00 | 2·62 |
| মোট | 153 | 100·00 | — | — | — | — |

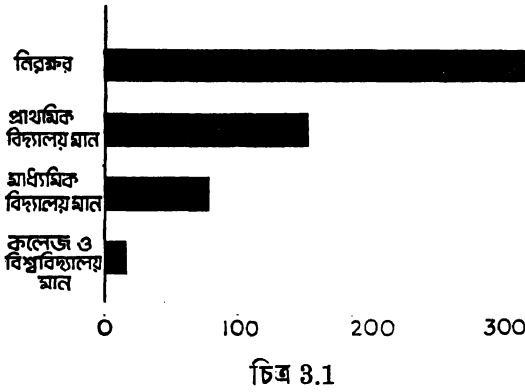
মনে রাখতে হবে, কোন শ্রেণীর ক্ষুদ্রতর-সূচক ক্রমযৌগিক পরিসংখ্যা বস্তুতপক্ষে শ্রেণীটির উর্ধ্বসীমান্তের ব'লে ধরতে হবে। তেমনি বৃহত্তর-সূচক ক্রম-যৌগিক পরিসংখ্যাটিকেও ধরতে হবে শ্রেণীটির অধঃসীমান্তের ব'লে। যেমন, 3.10 সারণী থেকে বলা যায়, 99·95 ডিঃ ফাঃ অথবা তার কম তাপমাত্রা ছিল মোট 121 দিন এবং 87·95 ডিঃ ফাঃ অথবা তার বেশী তাপমাত্রা ছিল মোট 145 দিন।

## 3.4 পরিসংখ্যা-বিভাজনের লৈখিক উপস্থাপন:

দ্বিতীয় পরিচ্ছেদে প্রকৃত মাপ-সূচক অর্থাৎ অ-পরিসংখ্যা রাশিতথ্যের লৈখিক উপস্থাপনের বিভিন্ন পদ্ধতিগুলি বিস্তারিতভাবে আলোচনা করা হয়েছে। বর্তমান অনুচ্ছেদে পরিসংখ্যা রাশিতথ্যের অর্থাৎ পরিসংখ্যা-বিভাজনের লৈখিক উপস্থাপনের প্রসঙ্গটি আলোচিত হবে।

## 3.4.1 গুণলক্ষণের পরিসংখ্যা-বিভাজনের লৈখিক উপস্থাপন :

কোন গুণলক্ষণের পরিসংখ্যা-বিভাজনের লৈখিক উপস্থাপনের জন্য পূর্ববর্তী পরিচ্ছেদে বর্ণিত স্তম্ভচিত্র এবং আপেক্ষিক পরিসংখ্যা-বিভাজন উপস্থাপনের জন্য খণ্ডিত স্তম্ভচিত্র অথবা বৃত্তচিত্র ব্যবহার করা যেতে পারে। 3.1 চিত্রটি 3.1 সারণীতে প্রদত্ত শিক্ষাগত মান অনুযায়ী গ্রামবাসীদের পরিসংখ্যা-বিভাজনের

চিত্র 3.1

468 জন গ্রামবাসীর শিক্ষাগত মানের স্তম্ভচিত্র ( সারণী 3.1 )।

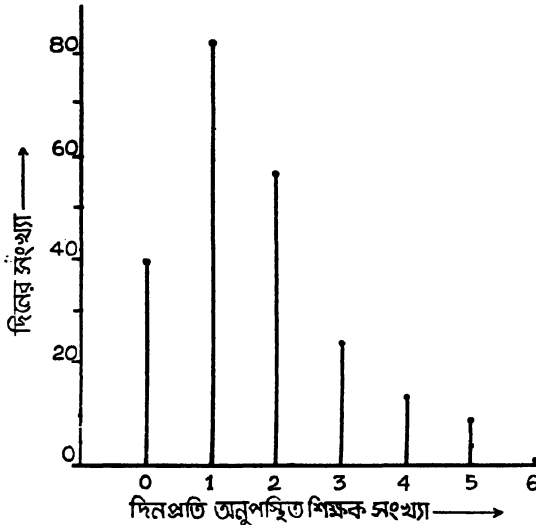
## 3.4.2 বিচ্ছিন্ন চলের পরিসংখ্যা-বিভাজনের লৈখিক উপস্থাপন :

সংশ্লিষ্ট লক্ষণটি একটি বিচ্ছিন্ন চল হলে পরিসংখ্যা-বিভাজনের লৈখিক উপস্থাপন নিম্নলিখিত যে কোন একটি পদ্ধতিতে হতে পারে :

**পরিসংখ্যা-স্তম্ভচিত্র** (column diagram) :

চলটির এক-একটি শ্রেণীতে এক-একটি মান নেওয়া হলে এই পদ্ধতিটি ব্যবহার করা যায়। সুবিধামত স্কেল ব্যবহারে অনুভূমিক অক্ষরেখায় বিভিন্ন বিন্দুর সাহায্যে চলের বিভিন্ন মান, এবং উলম্ব অক্ষটিতে পরিসংখ্যা নির্দেশ ক'রে চলের বিভিন্ন মান-নির্দেশী বিন্দু থেকে সংশ্লিষ্ট পরিসংখ্যার সমান দৈর্ঘ্যের লম্ব (column) উত্তোলন ক'রে পাওয়া যায় পরিসংখ্যা-স্তম্ভচিত্র। 3.2 চিত্রে 3.4 সারণীতে প্রদত্ত পরিসংখ্যা-বিভাজনের পরিসংখ্যা-স্তম্ভচিত্র দেওয়া হয়েছে।

**পরিসংখ্যা-বহুভুজ** ( frequency polygon) :

ধরা যাক, চলের মানকে ( শ্রেণী-অন্তরের ক্ষেত্রে শ্রেণী-মধ্যককে ) X-স্থানাঙ্ক এবং সংশ্লিষ্ট পরিসংখ্যাকে Y-স্থানাঙ্ক ধ'রে নিয়ে সুবিধামত স্কেল ব্যবহারে বিভিন্ন বিন্দুগুলি সংস্থাপন করা হ'ল। অতঃপর অনুভূমিক রেখার উপর

চিত্র 3.2

দিনপ্রতি অনুপস্থিত শিক্ষকসংখ্যার পরিসংখ্যা-স্তম্ভচিত্র ( সারণী 3.4 )।

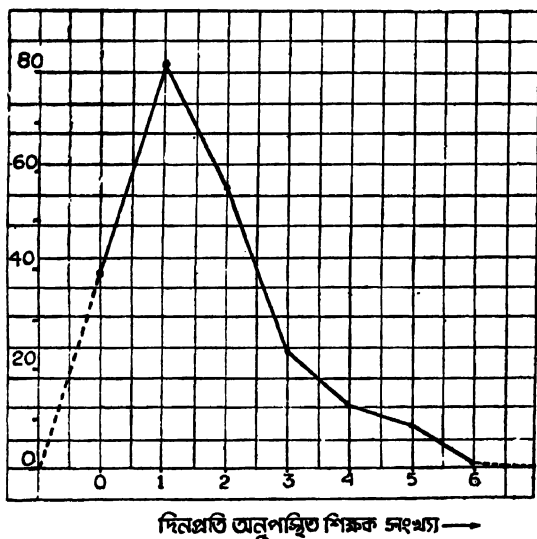
ক্ষুদ্রতম মানসূচক বিন্দুটির এক একক বামে একটি এবং বৃহত্তম মানসূচক বিন্দুটির এক একক দক্ষিণে একটি—মোট এই দুটি অতিরিক্ত বিন্দু নেওয়া হ'ল। এখন সন্নিহিত বিন্দুগুলি সরলরেখার সাহায্যে সংযুক্ত ক'রে অনুভূমিক অক্ষরেখার সহযোগে যে বহুভুজটি পাওয়া যায় সেইটিই পরিসংখ্যা-বহুভুজ।

3.3 চিত্রটিতে 3.4 সারণীতে প্রদত্ত পরিসংখ্যা-বিভাজনের পরিসংখ্যা-বহুভুজ দেখানো হয়েছে। লক্ষ্য কর, এখানে $(-1, 0)$ এবং $(7, 0)$ এই দুটি অতিরিক্ত বিন্দু সংস্থাপন করা হয়েছে।

### 3.4.3 অবিচ্ছিন্ন চলের পরিসংখ্যা-বিভাজনের লৈখিক উপস্থাপন :

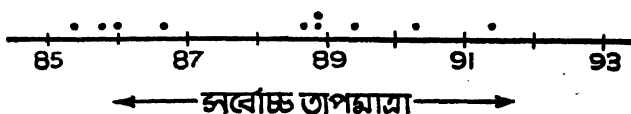
**বিন্দুচিত্র** (point-diagram) :

কোন অবিচ্ছিন্ন চল সম্পর্কে গৃহীত স্বল্পসংখ্যক মানের শ্রেণীবিন্যাস ছাড়াই লৈখিক উপস্থাপনা সম্ভব। 3.6 সারণীর প্রথম দশটি মান নেওয়া যাক।

প্রয়োজনীয় দৈর্ঘ্যের একটি সরলরেখার সাহায্যে একটি অবিচ্ছিন্ন তাপ মাপনামাত্রা (scale for measuring temperature) সূচিত ক'রে বিভিন্ন মানগুলিকে বিন্দু দ্বারা নির্দেশ ক'রে যে চিত্রটি পাওয়া যায় সেটি হ'ল বিন্দুচিত্র

চিত্র 3.3

দিনপ্রতি অনুপস্থিত শিক্ষকসংখ্যার পরিসংখ্যা-বহুভুজ ( সারণী 3.4 )।

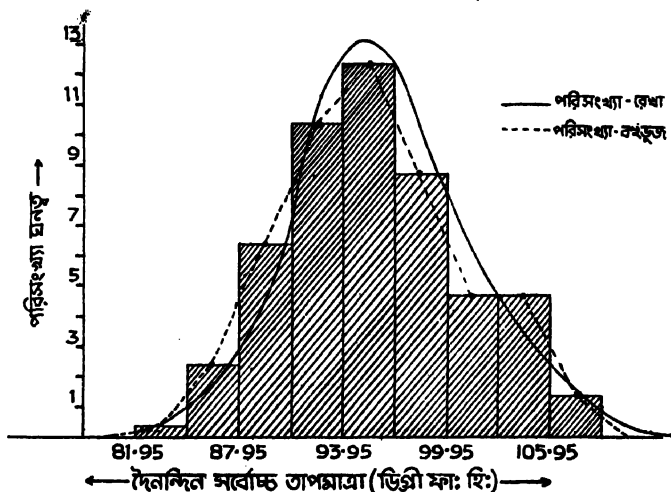
( চিত্র 3.4 )। বিচ্ছিন্ন চলের পরিসংখ্যা-বিভাজনও বিন্দুচিত্রের সাহায্যে উপস্থাপিত হতে পারে।

একই মান একাধিকবার পাওয়া গেলে একটির মাথায় আর একটি বিন্দু সংস্থাপন করা চলতে পারে। তবে মোট পরিসংখ্যার পরিমাণ অনেক বেশী হলে বিন্দুচিত্রটি অত্যন্ত দুর্বোধ্য হয়ে ওঠে। সেক্ষেত্রে বিভিন্ন মানের যথার্থ লেখচিত্র পাওয়ার আশা ত্যাগ ক'রে রাশিতথ্যগুলি প্রথমে শ্রেণীবিন্যস্ত করার পর শ্রেণীগুলির পরিসংখ্যা সংক্রান্ত লেখচিত্র নিয়েই সন্তুষ্ট থাকতে হয়।

চিত্র 3.4

কলকাতার ( 1—10 মার্চ, 1972) দৈনন্দিন সর্বোচ্চ তাপমাত্রার বিন্দুচিত্র ( সারণী 3.6 )।

শ্রেণী-অন্তরগুলি সমদৈর্ঘ্য হলে পরিসংখ্যা-বহুভুজ ব্যবহার করা যেতে পারে এক্ষেত্রেও। এখানে পরিসংখ্যাগুলি শ্রেণী-মধ্যকের পরিসংখ্যা ব'লে ধ'রে নেওয়া হয় ( চিত্র 3.5 )।

**আয়তচিত্র** (histogram) **:**

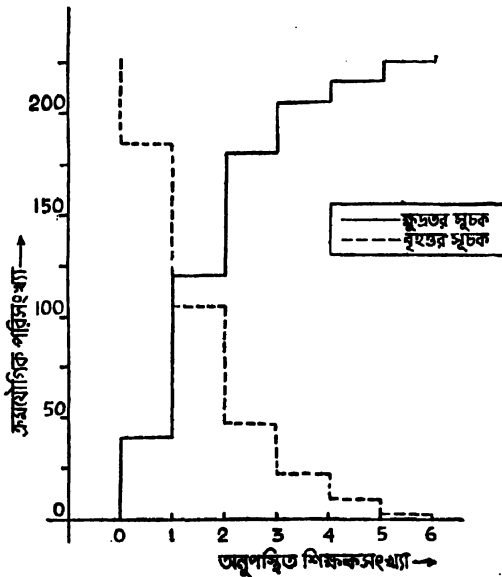
অবিচ্ছিন্ন চল সংক্রান্ত পরিসংখ্যা-বিভাজনের লৈখিক উপস্থাপনের আদর্শ পদ্ধতি হ'ল আয়তচিত্রের ব্যবহার। এখানে সুবিধামত স্কেল ব্যবহারে অনুভূমিক অক্ষটিতে বিভিন্ন শ্রেণী-সীমান্তগুলি নির্দেশ করা হয়, আর উল্লম্ব অক্ষটিতে নেওয়া হয় পরিসংখ্যা-ঘনত্ব। অতঃপর বিভিন্ন শ্রেণী-অন্তরের ওপর সংশ্লিষ্ট পরিসংখ্যা-ঘনত্বের সমান উচ্চতা সম্পন্ন এক-একটি আয়তক্ষেত্র আঁকা হয়। স্পষ্টতঃই আয়তক্ষেত্রগুলির ক্ষেত্রফল সংশ্লিষ্ট শ্রেণী-পরিসংখ্যার সমানুপাতী; সুতরাং শ্রেণী-অন্তরগুলি অসমদৈর্ঘ্য হলেও আয়তচিত্রে এগুলি পরস্পর তুলনীয় হয়। $i$-তম শ্রেণী-অন্তরের দৈর্ঘ্য $w_i$, এবং পরিসংখ্যা $f_i$, সংশ্লিষ্ট আয়তক্ষেত্রের ক্ষেত্রফল $A_i$ এবং উচ্চতা $h_i$ দ্বারা সূচিত হলে, গাণিতিকভাবে $f_i \propto A_i$, অর্থাৎ $f_i \propto w_i h_i$ বা, $f_i \propto h_i$ যদি $w_i$ ধ্রুবক হয়, অর্থাৎ শ্রেণীগুলি সমদৈর্ঘ্য হয়। সুতরাং দেখা যাচ্ছে, সমদৈর্ঘ্য শ্রেণীবিন্যাসের ক্ষেত্রে আয়তচিত্রের উল্লম্ব অক্ষরেখায় পরিসংখ্যা-ঘনত্বের পরিবর্তে শুধু পরিসংখ্যাও নেওয়া যেতে

চিত্র 3.5

কলকাতার ( মার্চ—জুলাই, 1972) দৈনন্দিন সর্বোচ্চ তাপমাত্রার পরিসংখ্যা-বিভাজনের ( সারণী 3.8) আয়তচিত্র ও পরিসংখ্যা-বহুভুজ এবং পরিসংখ্যা-রেখা।

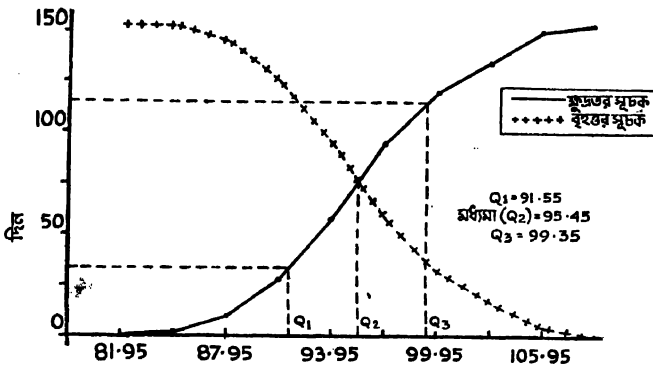
পারে। তবে যে কোন ধরনের শ্রেণীবিন্যাসের ক্ষেত্রেই পরিসংখ্যা-ঘনত্বের ব্যবহারই প্রশস্ত।

আয়তচিত্রের অনুভূমিক অক্ষরেখায় শ্রেণী-সীমান্তগুলি নির্দেশ করায় চিত্রে অঙ্কিত আয়তচিত্রগুলি থাকে পরস্পর সন্নিবদ্ধ ( চিত্র 3.5 দ্রষ্টব্য )।

## 3.5 ক্রমযৌগিক পরিসংখ্যা-বিভাজনের লৈখিক উপস্থাপন :

প্রথমে বিচ্ছিন্ন চলের প্রসঙ্গে আসা যাক। চলটির সংশ্লিষ্ট মানগুলির বিপরীতে ক্রমযৌগিক ( ক্ষুদ্রতর বা বৃহত্তর ) পরিসংখ্যা-সূচক বিন্দুগুলি সংস্থাপন করা হয়। এখানে পরস্পর সন্নিহিত দুটি মানের মধ্যে চলটির অন্য মান থাকা সম্ভব নয়, সুতরাং দুটি মানের মধ্যবর্তী স্থানে ক্রমযৌগিক পরিসংখ্যাও পরিবর্তিত হয় না। তাই এক্ষেত্রে **ক্রমযৌগিক পরিসংখ্যা-রেখা** (ogive) পাওয়ার জন্য সন্নিহিত বিন্দুগুলি **সোপানচিত্রের** (step diagram) আকারে পরস্পর যুক্ত করা হয় ( চিত্র 3.6 )। ক্ষুদ্রতর- এবং বৃহত্তর-সূচক ক্রমযৌগিক পরিসংখ্যা-নির্দেশী দুটি সোপানচিত্র পাওয়া যেতে পারে।

চিত্র 3.6

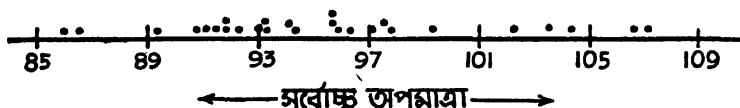
দিনপ্রতি অনুপস্থিত শিক্ষকসংখ্যার ক্রমযৌগিক পরিসংখ্যা-বিভাজনের সোপানচিত্র ( সারণী 3.5 )।

অবিচ্ছিন্ন চলের ক্ষেত্রে ক্ষুদ্রতর- এবং বৃহত্তর-সূচক পরিসংখ্যাগুলি যথাক্রমে শ্রেণীগুলির উর্ধ্বসীমান্ত এবং অধঃসীমান্তের বিপরীতে সংস্থাপন করা হয়। ক্ষুদ্রতর-সূচক ক্রমযৌগিক পরিসংখ্যা-রেখার জন্য প্রথম শ্রেণীর অধঃসীমান্ত-সূচক বিন্দুটি এবং বৃহত্তর-সূচক ক্রমযৌগিক পরিসংখ্যা-রেখার জন্য শেষশ্রেণীর উর্ধ্বসীমান্ত-সূচক বিন্দুটি অতিরিক্ত নেওয়া হয় অনুভূমিক রেখার ওপরে। এক্ষেত্রে একই শ্রেণীর অধঃ- এবং উর্ধ্বসীমান্তের মধ্যে ক্রমযৌগিক পরিসংখ্যা ক্রমবর্ধমান বা ক্রমহ্রাসমান মনে করা যেতে পারে, তাই এক্ষেত্রে ক্রমযৌগিক পরিসংখ্যা-রেখা পাওয়া যায় সন্নিহিত বিন্দুগুলি পরস্পর সাধারণভাবে সরলরেখার সাহায্যে যোগ ক'রে। 3.7 চিত্রে 3.10 সারণীতে প্রদত্ত ক্রমযৌগিক পরিসংখ্যা-বিভাজনের লেখচিত্র আঁকা হয়েছে।

চিত্র 3.7

কলকাতার ( মার্চ—জুলাই, 1972) দৈনন্দিন সর্বোচ্চ তাপমাত্রার ক্রমযৌগিক পরিসংখ্যা-রেখা চিত্র ( সারণী 3.10 )।

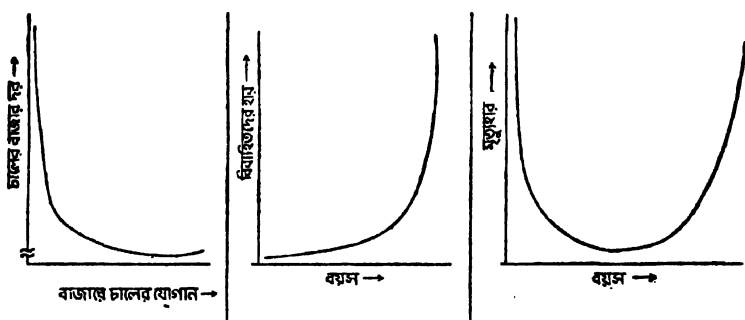
## 3.6 পরিসংখ্যা-রেখা :

3.4 চিত্রের বিন্দুচিত্রটিতে দৈনন্দিন সর্বোচ্চ তাপমাত্রা এই চলটির 10টি মান উপস্থাপিত হয়েছে। এই চিত্রটি থেকে আমরা চলটির প্রকৃতি সম্বন্ধে তেমন কিছু জানতে পারি না। কিন্তু চিত্রটিতে ক্রমশঃ আরও বেশী সংখ্যক মান নেওয়া হতে থাকলে চলটির পরিসংখ্যা-বিভাজনের বিভিন্ন বৈশিষ্ট্য সম্বন্ধে ক্রমশঃ একটি স্পষ্টতর চিত্র ফুটে উঠবে। যেমন, 30টি মান সম্বলিত ( 3.6 সারণীর প্রথমটি থেকে শুরু ক'রে প্রতি পঞ্চম মানটি নিয়ে ) 3.8 বিন্দুচিত্রটি থেকে দেখা যাচ্ছে গৃহীত মানগুলির 95·0 এর কাছাকাছি অবস্থান করার দিকে প্রবণতা রয়েছে।

অবশ্য মোট পরিসংখ্যা অনেক বেশী হলে পরিসংখ্যা রাশিতথ্য আয়তচিত্রে উপস্থাপিত করা হয়। আয়তচিত্রেই চলটির পরিসংখ্যা-বিভাজনের উল্লিখিত বৈশিষ্ট্যগুলি চোখে পড়বে। এখন মোট পরিসংখ্যা ক্রমাগত বাড়িয়ে গেলে এবং সেইসঙ্গে শ্রেণীদৈর্ঘ্যগুলিও ক্রমাগত কমিয়ে আনলে ( অর্থাৎ শ্রেণীসংখ্যা বাড়িয়ে গেলে ) পরিসংখ্যা-বিভাজনের অন্তর্নিহিত ধাঁচটি ক্রমশঃ স্পষ্টতর হয়ে ফুটে উঠবে।

চিত্র 3.8

**কলকাতার দৈনন্দিন সর্বোচ্চ তাপমাত্রার বিন্দুচিত্র ( 3.6 সারণীর প্রথমটি থেকে শুরু ক'রে প্রতি পঞ্চম মান )।**

এইভাবে বিন্দুচিত্রে মোট পরিসংখ্যা ক্রমাগত বর্ধিত ক'রে বিন্দুগুলির বিন্যাসের,—অথবা আয়তচিত্রে যুগপৎ মোট পরিসংখ্যা এবং শ্রেণীসংখ্যা ক্রমাগত বর্ধিত ক'রে আয়তশীর্ষের মধ্যবিন্দুগুলির বিন্যাসের যে ক্রমাসন্ন রেখাচিত্রটি পাওয়া যায়, সেটিকে বলা হয় **পরিসংখ্যা-রেখা** (frequency curve) ( চিত্র 3.5 )। আলোচ্য ক্ষেত্রে পরিসংখ্যা-রেখাটি টুপির বা ঘণ্টার আকৃতিবিশিষ্ট (bell-shaped)। অধিকাংশ অবিচ্ছিন্ন চলের পরিসংখ্যা-রেখাই ঘণ্টাকৃতি। অবশ্য অন্য আকৃতির পরিসংখ্যা-রেখাও দেখা যায়। যেমন উল্টো J- আকৃতিবিশিষ্ট ( চিত্র 3.9*a* ), J-আকৃতিবিশিষ্ট ( চিত্র 3.9*b* ) ও U-আকৃতিবিশিষ্ট ( চিত্র 3.9*c* ), ইত্যাদি।

আগেই বলা হয়েছে, রাশিবিজ্ঞানে বিশ্লেষণের প্রয়োজনে যে ধরনের রাশিতথ্য

চিত্র 3.9

**(a) উল্টো J-আকৃতি, (b) J-আকৃতি এবং (c) U-আকৃতিবিশিষ্ট পরিসংখ্যা-রেখা।**

নিয়ে সাধারণতঃ নাড়াচাড়া করা হয় সেগুলি অধিকাংশ ক্ষেত্রেই একটি বৃহত্তর সমষ্টির, যাকে আমরা সমগ্রক (universe) বলি, তার নমুনা (sample) বিশেষ। যদি সমগ্রকের অন্তর্গত প্রতিটি তথ্যই আমাদের বিশ্লেষণের আওতায় আসত তাহলে আমরা পরিসংখ্যা-বিভাজন নিরূপণ করার জন্য খুব কম দৈর্ঘ্যের শ্রেণী নিতে পারতাম। সেক্ষেত্রে স্পষ্টতঃই পরিসংখ্যা-বহুভুজটির ক্রমাসন্ন রূপই হতো পরিসংখ্যা-রেখা। এই কারণে নমুনালব্ধ তথ্য থেকে সমগ্রকে চলটির বিভাজনের প্রকৃতি এবং রূপ সম্বন্ধে কিছুটা ধারণা পাওয়ার জন্য পরিসংখ্যা-বহুভুজের ধাঁচটি অনুসরণ ক'রে পরিসংখ্যা-রেখাটি এঁকে নেওয়া হয়। পরিসংখ্যা-বহুভুজের মূল ধাঁচটি বজায় রেখে একটি মসৃণ রেখাই আঁকা হয় এক্ষেত্রে—বহুভুজটির প্রতিটি শীর্ষবিন্দুর মধ্য দিয়েই এটিকে যেতে হবে এমন কোন কথা নেই।

অনুরূপভাবে ক্রমযৌগিক পরিসংখ্যা-বিভাজনের জন্যও মসৃণ ক্রমযৌগিক পরিসংখ্যা-রেখা আঁকা যায়।

## 3.7 অনুশীলনী

3.1 রাশিতথ্য সংক্ষেপীকরণের প্রয়োজন হয় কেন? পরিসংখ্যা-বিভাজনের সাহায্যে কি-ভাবে বিভিন্ন ধরনের রাশিতথ্য সংক্ষেপ করা যায় বর্ণনা কর।

3.2 গুণলক্ষণ এবং চলের মধ্যে পার্থক্য নির্দেশ কর। বিচ্ছিন্ন চল এবং অবিচ্ছিন্ন চলের সংজ্ঞা দাও। উপযুক্ত উদাহরণ সহযোগে এদের পার্থক্য নির্দেশ কর। নীচের উদাহরণগুলিতে কোন্‌টি গুণলক্ষণ এবং কোন্‌টি চল নির্দেশ কর। চলের ক্ষেত্রে কোন্‌টি বিচ্ছিন্ন এবং কোন্‌টি অবিচ্ছিন্ন চল বল: (i) ছাত্রের বয়স, (ii) ছাত্রের গত জন্মদিনে বয়স, (iii) নির্দিষ্ট পরিমান চালের দাম, (iv) কলকাতা বাজারে একদিনের চালকেনার খরচ, (v) শিক্ষাগত যোগ্যতা, (vi) ব্যক্তিগত মালিকানায় জমির পরিমাণ, (vii) পরীক্ষায় কৃতকার্যতা, (viii) পরীক্ষায় প্রাপ্ত নম্বর, (ix) পরিবারের আয়তন (সদস্যসংখ্যা), (x) জমির আয়তন।

3.3 উদাহরণসহ সংজ্ঞা দাও: (i) পরিসংখ্যা, (ii) আপেক্ষিক পরিসংখ্যা, (iii) পরিসংখ্যা-বিভাজন, (iv) পরিসংখ্যা-ঘনত্ব, (v) শ্রেণী-দৈর্ঘ্য, (vi) শ্রেণী-সীমা, (vii) শ্রেণী-সীমান্ত, (viii) শ্রেণী-মধ্যক, (ix) ক্রমযৌগিক পরিসংখ্যা, (x) ক্রমযৌগিক আপেক্ষিক পরিসংখ্যা।

3.4 অবিচ্ছিন্ন চল সংক্রান্ত অবিন্যস্ত রাশিতথ্য থেকে পরিসংখ্যা-বিভাজন

গঠন করতে হলে কোন্ কোন্ বিষয়ে লক্ষ্য রাখতে হবে? উদাহরণসহ ব্যাখ্যা কর।

3.5 পরিসংখ্যা-বিভাজনের লৈখিক উপস্থাপনের পদ্ধতিগুলি বর্ণনা কর। প্রমাণ কর যে, সমদৈর্ঘ্য শ্রেণীবিন্যাসের ক্ষেত্রে লব্ধ পরিসংখ্যা-বহুভুজ এবং আয়তলেখের আয়তন অভিন্ন।

3.6 পরিসংখ্যা-রেখার সংজ্ঞা দাও। 3.8 অনুশীলনীতে পরিসংখ্যা-রেখাটি অঙ্কন কর। নিম্নলিখিত উদাহরণগুলিতে সাধারণতঃ কোন্ আকৃতির পরিসংখ্যা-রেখা পাওয়া সম্ভব আলোচনা কর :

(i) কোন বিদ্যালয়ের ছাত্রদের উচ্চতা অনুযায়ী বিভাজন; (ii) কোন শহরের অধিবাসীদের মাসিক আয় অনুযায়ী বিভাজন; (iii) কোন দেশের অধিবাসীদের বয়স অনুযায়ী মৃত্যুহার; (iv) এক বছরে ঘটানো দুর্ঘটনার সংখ্যা অনুযায়ী গাড়ীচালকদের বিভাজন; (v) কোন দেশে 60 বৎসর এবং তদূর্ধ্ব বয়সী অধিবাসীদের বয়স অনুযায়ী মৃত্যুহার; (vi) পরীক্ষায় প্রাপ্ত নম্বর অনুযায়ী ছাত্রদের বিভাজন।

3.7 1974 সালে কলকাতা সিনিয়র ডিভিশন ফুটবল লীগে বিভিন্ন খেলায় গোলের মোট সংখ্যা নীচে দেওয়া হ'ল :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 4 | 4 | 2 | 2 | 2 | 3 | 0 | 0 | 3 |
| 2 | 5 | 3 | 2 | 3 | 4 | 2 | 2 | 3 | 2 |
| 2 | 2 | 1 | 0 | 3 | 0 | 5 | 0 | 2 | 1 |
| 3 | 2 | 0 | 6 | 0 | 5 | 0 | 2 | 4 | 0 |
| 1 | 2 | 0 | 1 | 3 | 2 | 2 | 2 | 0 | 0 |
| 3 | 1 | 4 | 0 | 2 | 6 | 0 | 1 | 0 | 0 |
| 5 | 4 | 3 | 1 | 0 | 0 | 0 | 1 | 0 | 2 |
| 1 | 0 | 0 | 1 | 0 | 1 | 1 | 2 | 1 | 1 |
| 2 | 2 | 4 | 4 | 2 | 2 | 3 | 3 | 0 | 4 |
| 2 | 0 | 0 | 1 | 0 | 2 | 2 | 3 | 4 | 0 |
| 0 | 5 | 1 | 0 | 1 | 3 | 2 | 0 | 0 | 2 |
| 0 | 2 | 0 | 4 | 2 | 2 | 2 | 1 | 0 | 3 |
| 0 | 0 | 1 | 2 | 1 | 0 | 1 | 1 | 1 | 0 |
| 2 | 2 | 2 | 5 | 0 | 3 | 1 | 1 | 2 | 4 |
| 1 | 1 | 5 | 4 | 2 | 4 | 3 | 1 | 1 | 4 |
| 0 | 1 | 2 | 4 | 3 | 1 | 2 | 1 | 1 | 1 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 2 | 0 | 2 | 0 | 1 | 0 | 1 | 1 | 1 |
| 2 | 0 | 0 | 1 | 2 | 1 | 0 | 5 | 1 | 1 |
| 0 | 1 | 1 | 0 | 0 | 3 | 2 | 1 | 2 | 1 |

প্রদত্ত রাশিতথ্য থেকে গোলসংখ্যার পরিসংখ্যা-বিভাজন গঠন কর। পরিসংখ্যা-বিভাজন সারণী থেকে (i) ঠিক 2 খানি, (ii) বড় জোর 2 খানি এবং (iii) অন্ততঃ 2 খানি গোল হয়েছে এমন খেলার সংখ্যা, অনুপাত এবং শতকরা হার নির্ণয় কর। পরিসংখ্যা-বিভাজনটি উপযুক্ত লেখচিত্রের সাহায্যে উপস্থাপিত কর।

3.8 ইছাপুর হাই স্কুলের 1970 সালের বাৎসরিক পরীক্ষায় পঞ্চম শ্রেণীর 102 জন ছাত্রের অঙ্কের নম্বর নীচে দেওয়া হ'ল। সমান দৈর্ঘ্যের 10টি শ্রেণী নিয়ে একটি পরিসংখ্যা-বিভাজন সারণী গঠন কর। সারণীতে গৃহীত শ্রেণীগুলির সীমা, সীমান্ত, মধ্যক, ক্রমযৌগিক পরিসংখ্যা এবং পরিসংখ্যা-ঘনত্ব দেখাও। বিভাজনটি আয়তচিত্র এবং পরিসংখ্যা-বহুভুজের সাহায্যে উপস্থাপিত কর এবং সংশ্লিষ্ট ক্রমযৌগিক পরিসংখ্যা-রেখা দুটি অঙ্কিত কর। শেষোক্ত চিত্র-দুটি থেকে উত্তীর্ণ ছাত্রদের আনুমানিক সংখ্যা ( উত্তীর্ণ হওয়ার জন্য অন্যূন 34% নম্বর পাওয়া প্রয়োজন ) এবং দ্বিতীয় বিভাগে উত্তীর্ণদের ( দ্বিতীয় বিভাগে উত্তীর্ণ হওয়ার জন্য 45% থেকে 59% নম্বর পেতে হয় ) আনুমানিক সংখ্যা নির্ণয় কর।

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 10 | 25 | 56 | 49 | 16 | 26 | 72 | 23 |
| 38 | 53 | 33 | 54 | 30 | 62 | 9 | 74 |
| 71 | 98 | 70 | 34 | 42 | 76 | 38 | 41 |
| 54 | 44 | 30 | 30 | 20 | 48 | 41 | 21 |
| 74 | 40 | 4 | 39 | 38 | 36 | 11 | 30 |
| 37 | 99 | 8 | 34 | 31 | 32 | 30 | 30 |
| 30 | 9 | 5 | 4 | 33 | 43 | 47 | 32 |
| 20 | 38 | 37 | 16 | 15 | 16 | 11 | 6 |
| 30 | 14 | 16 | 43 | 31 | 30 | 32 | 18 |
| 55 | 23 | 17 | 35 | 18 | 20 | 30 | 10 |
| 43 | 17 | 33 | 30 | 74 | 87 | 15 | 62 |
| 34 | 33 | 5 | 18 | 12 | 7 | 31 | 17 |
| 40 | 9 | 28 | 15 | 30 | 30 | | |

## 3.8 নির্দেশিকা

1. Goon, A. M., Gupta, M. K. & Dasgupta, B. *Fundamentals of Statistics* (Vol I). World Press, 1975.

2. Mills, F. C. *Statistical Methods*. H. Holt, 1955.

3. Moore, P. G. *Principles of Statistical Techniques*. Cambridge University Press, 1969.

4. Wallis, A. W. & Roberts, H. V. *Statistics : a New Approach*. Methuen, 1957.

5. Yule, G. U. & Kendall, M. G. *An Introduction to The Theory of Statistics*. Charles Griffin, 1955.

# 4 মধ্যগামিতা এবং মধ্যগামিতা-মাপক
## ( Central tendency and its measures )

### 4.1 বিবরণাত্মক মাপকাবলী (descriptive measures) ঃ

সারণীবিন্যাস, পরিসংখ্যা-বিভাজন নিরূপণ, লৈখিক উপস্থাপন, প্রভৃতি হ'ল রাশিতথ্য সংক্ষেপীকরণের প্রথম ধাপ। কিন্তু রাশিতথ্য সংগ্রহের মূল উদ্দেশ্যের সবটা প্রায়ই এই পর্যায়ে সাধিত হয় না—বিশেষ ক'রে পরিসংখ্যা-সূচক রাশিতথ্য সম্পর্কে আমাদের আরও বিশ্লেষণের প্রয়োজন হয়। এই সব ক্ষেত্রে সাধারণতঃ আমাদের আসল উদ্দেশ্য থাকে চলটির বিভাজনের এক বা একাধিক বৈশিষ্ট্যের ওপর আলোকপাত করা। যেমন, কোন একটি বিশ্ববিদ্যালয়-পরীক্ষায় পরীক্ষার্থী-দের প্রাপ্ত নম্বর-সংক্রান্ত রাশিতথ্য সংগ্রহের উদ্দেশ্য হতে পারে কতজন উত্তীর্ণ হয়েছে, উত্তীর্ণদের কতজন প্রথম বিভাগে আছে, সর্বোচ্চ নম্বর কত, সর্বনিম্নই বা কত, গড়ে কী রকম নম্বর উঠেছে, সাধারণভাবে পরীক্ষার্থীদের পরস্পরের মধ্যে প্রাপ্ত নম্বরে কী ধরনের পার্থক্য রয়েছে—ইত্যাদি প্রশ্নের উত্তর জানা। স্পষ্টতঃই কোনও চলের বিভাজন-সংক্রান্ত এই ধরনের বিভিন্ন বৈশিষ্ট্য প্রকাশ করা যেতে পারে এক-একটি একক সংখ্যার সাহায্যে। এইসব সংখ্যার সাহায্যে চলের বিভিন্ন বৈশিষ্ট্যের বিবরণ পাওয়া যায়—তাই এদের বলা হয় **বিবরণাত্মক মাপক**। একই ধরনের একাধিক চলের পরিসংখ্যা-বিভাজনের তুলনা করা হয়ে থাকে এইসব বিবরণাত্মক মাপকের সাহায্যে। যেমন, 1970 সালে পি.ইউ. পরীক্ষায় প্রেসিডেন্সি কলেজের এবং বেলুড় বিদ্যামন্দিরের ছাত্র-ছাত্রীদের ফলাফল তুলনা করতে হলে দুটি কলেজের ছাত্র-ছাত্রীদের ঐ পরীক্ষায় প্রাপ্ত নম্বরের গড়, প্রথম শ্রেণীতে উত্তীর্ণদের হার, ফেলের হার, ইত্যাদি তুলনা করা হবে। লক্ষণীয়, এক অর্থে পরিসংখ্যা-বিভাজন সারণীর অন্তর্গত প্রতিটি পরিসংখ্যাই এক-একটি বিবরণাত্মক মাপক।

বর্তমান এবং পরবর্তী দুটি পরিচ্ছেদে বিভিন্ন বিবরণাত্মক মাপক সম্পর্কে বিস্তারিত আলোচনা করা হয়েছে।

### 4.2 মধ্যগামিতা (central tendency) ঃ

সারণী 3.4 এবং সারণী 3.7-এ উপস্থাপিত মিলচিহ্নগুলি কিংবা সংশ্লিষ্ট বিন্দুচিত্র ( চিত্র 3.4 ) বা আয়তচিত্রের ( চিত্র 3.5 ) দিকে তাকালে অথবা সংশ্লিষ্ট পরিসংখ্যা-বিভাজন সারণী-দুটি লক্ষ্য করলে দেখা যাবে, বিচ্ছিন্ন বা অবিচ্ছিন্ন

চল-সংক্রান্ত দুটি উদাহরণেই সংগৃহীত বিভিন্ন মানগুলির চলের মানসীমার মাঝামাঝি অবস্থিত বিশেষ একটি মানের কাছাকাছি গুচ্ছবদ্ধ হওয়ার প্রবণতা দেখা যাচ্ছে—এই বিশেষ মানটি থেকে উভয় দিকে যত দূরে যাওয়া যায়, দেখা যাবে এই প্রবণতা ততই কমের দিকে। সংগৃহীত মানগুলির মধ্যবর্তী কোন মানের কাছাকাছি গুচ্ছবদ্ধ হওয়ার এই প্রবণতা চলের পরিসংখ্যা-বিভাজনের একটি লক্ষণীয় এবং গুরুত্বপূর্ণ বৈশিষ্ট্য। রাশিবিজ্ঞানের পরিভাষায় এই প্রবণতাকে বলা হয় চলটির ( বা সংশ্লিষ্ট পরিসংখ্যা-বিভাজনের ) **মধ্যগামিতা**। মোট পরিসংখ্যার পরিমাণ মোটামুটি বেশী হলে চলের পরিসংখ্যা-বিভাজন থেকে এই বৈশিষ্ট্যটি সহজেই চোখে পড়ে। মোট পরিসংখ্যা কম হলেও ভালভাবে লক্ষ্য করলে বৈশিষ্ট্যটি টের পাওয়া যায়। 4.1 সারণীটি লক্ষ্য করলে চোখে পড়বে, একর-প্রতি ফলনের হার 8·5 কুইণ্টালের নিকটবর্তী কোন মানের দিকে গুচ্ছবদ্ধ।

## সারণী 4.1

### 10 খণ্ড জমিতে একর-প্রতি ধানের ফলনের হার

| ভূমিখণ্ডের ক্রমিক সংখ্যা | একর-প্রতি ফলন ( কুইণ্টালে ) |
|---|---|
| 1 | 7·6 |
| 2 | 9·1 |
| 3 | 8·6 |
| 4 | 9·0 |
| 5 | 8·5 |
| 6 | 7·2 |
| 7 | 9·5 |
| 8 | 8·2 |
| 9 | 8·3 |
| 10 | 8·4 |
| মোট | 84·4 |

মধ্যগামিতার বিচারে এইভাবে একটি বিশেষ মানের সন্ধান পাওয়া গেলে মানটিকে প্রয়োজনবোধে সংগৃহীত রাশিতথ্যের প্রতিভূ হিসারে ব্যবহার করা চলে। অবিন্যস্ত অথবা গোষ্ঠীবদ্ধ রাশিতথ্য থেকে মধ্যগামিতা-নির্দেশক এই ধরনের একটি কেন্দ্রীয় মান বের করার জন্য সংগৃহীত মানগুলির ওপর বিশেষ বিশেষ গাণিতিক প্রক্রিয়া প্রয়োগ করা হয়। এইভাবে সংগৃহীত মানগুলির যে বিশেষ গাণিতিক প্রকাশন (mathematical expressions) পাওয়া যায় সেগুলিকে বলা হয় **মধ্যগামিতা-মাপক** (measures of central tendency)। মধ্যগামিতা বিভিন্ন দৃষ্টিকোণ থেকে পরিমাপ করা যেতে পারে, তাই রাশি-বিজ্ঞানে একাধিক মধ্যগামিতা-মাপকের ব্যবহার রয়েছে। বর্তমান পরিচ্ছেদে এগুলির মধ্যে প্রধান তিনটি, যথা **গাণিতিক গড়** (arithmetic mean), **মধ্যমা** বা **মধ্যমমান** (median) এবং **ভূয়িষ্ঠক** (mode) সম্বন্ধে বিস্তারিত আলোচনা করা হয়েছে। অপেক্ষাকৃত কম প্রচলিত মাপকগুলি সম্বন্ধে সংক্ষেপে উল্লেখ করা হয়েছে পরিচ্ছেদের শেষের দিকে।

মধ্যগামিতা মাপকের মান পরিসংখ্যা-বিভাজনের **অবস্থিতি** (location) কিছুটা নির্দেশ করে, এই জন্য একে অনেক সময় **অবস্থিতি-মাপকও** (measure of location) বলা হয়ে থাকে।

## 4.3 গাণিতিক গড় :

**4.3.1 সংজ্ঞা :** বিদ্যালয়পাঠ্য গণিতেই তোমরা গাণিতিক গড়ের সঙ্গে পরিচিত হয়েছ। প্রদত্ত মানগুলির সমষ্টিকে মানগুলির সংখ্যা দিয়ে ভাগ ক'রে পাওয়া যায় গাণিতিক গড় ( বা সংক্ষেপে, গড় )। প্রদত্ত মানগুলি যে এককে (unit) প্রদত্ত গাণিতিক গড়ের এককও তাই হবে।

মনে কর, $X$ চলটির $x_1, x_2, \ldots, x_n$—এই $n$টি মান দেওয়া আছে। চলটির প্রদত্ত মানগুলির গাণিতিক গড় $\bar{x}$ দ্বারা সূচিত হলে

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i. \qquad \cdots \qquad (4.1)$$

[ $\Sigma$ (sigma ; উচ্চারণ : সিগ্‌মা ) চিহ্নটিকে যোগচিহ্ন বলে।

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n \text{ ]}$$

কয়েকটি অবিন্যস্ত মানের পরিবর্তে চলটির পরিসংখ্যা-বিভাজন দেওয়া থাকতে পারে। বিচ্ছিন্ন চলের ক্ষেত্রে এক-একটি মানকে এক-একটি শ্রেণী হিসাবে ধরা হলে বিভিন্ন মানগুলি $x_i$ এবং সংশ্লিষ্ট পরিসংখ্যাগুলি $f_i$, $i=1(1)k$, দ্বারা চিহ্নিত করা যেতে পারে। এক্ষেত্রে (4.1) সূত্রটির সুসম্বদ্ধ রূপ হবে :

$$\bar{x}=\sum_{i=1}^{k} x_i f_i \Big/ \sum_{i=1}^{k} f_i$$

$$=\frac{1}{n}\sum_{i=1}^{k} x_i f_i. \qquad \cdots \qquad (4.2)$$

এখানে $n=\sum_{i=1}^{k} f_i$, অর্থাৎ মোট পরিসংখ্যা।

পরিসংখ্যা-বিভাজন নিরূপণ করার সময় এক-একটি শ্রেণীতে একাধিক মান গৃহীত হলে, আগেই বলা হয়েছে পরিসংখ্যা-বিভাজন সারণী থেকে বিভিন্ন ব্যষ্টির যথার্থ মান স্বতন্ত্রভাবে পাওয়া সম্ভব হয় না। সুতরাং এ থেকে বিভাজনটির গাণিতিক গড়ের যথার্থ মানও পাওয়া সম্ভব নয়। অবশ্য বিভিন্ন শ্রেণী-মধ্যকগুলিকে সংশ্লিষ্ট শ্রেণীর প্রতিনিধিস্থানীয় মান ধ'রে নিয়ে এবং সংশ্লিষ্ট পরিসংখ্যাকে শ্রেণী-মধ্যকের পরিসংখ্যা হিসেবে গণ্য ক'রে, (4.2) সূত্রের সাহায্যে গাণিতিক গড়ের একটি আসন্ন মান পাওয়া সম্ভব। শ্রেণীদৈর্ঘ্যগুলি মূল প্রসারের তুলনায় বেশ কম হলে এই আসন্ন মানে ভ্রান্তির পরিমাণ মোটামুটিভাবে উপেক্ষণীয় হয়।

আমাদের পরবর্তী আলোচনায় যেখানে যেখানে শ্রেণী-মধ্যককে শ্রেণী-প্রতিভূ ধ'রে নেওয়া হয়েছে রাশিতথ্য বিশ্লেষণের প্রয়োজনে, সেই সমস্ত ক্ষেত্রেই আমাদের উপরোক্ত মন্তব্য প্রযোজ্য হবে। এইসব ক্ষেত্রে ( যেমন 4.2 সূত্রে ) $x_i$-কে সাধারণভাবে $i$-তম শ্রেণীর যথার্থ মান অথবা প্রতিভূমান ( যেখানে যেমন ) বলা হবে।

**উদা. 4.1** 4.1 সারণীতে প্রদত্ত রাশিতথ্য থেকে একর-প্রতি ফলনের হারের গড় মান হবে

$$\bar{x}=\frac{7{\cdot}6+9{\cdot}1+\cdots+8{\cdot}4}{10} \text{ কুইণ্টাল} = 8{\cdot}44 \text{ কুইণ্টাল।}$$

**উদা. 4.2** 3.4 সারণীতে প্রদত্ত রাশিতথ্যের গাণিতিক গড় নির্ণয় করার জন্য নিম্নলিখিত ছকে অঙ্কপাতন করতে হবে।

**সারণী 4.2**

**বিদ্যালয়ে দিনপ্রতি অনুপস্থিত শিক্ষকসংখ্যার গড় নির্ণয়**

| অনুপস্থিত শিক্ষকদের সংখ্যা $x_i$ | দিনের সংখ্যা $f_i$ | $x_i f_i$ |
|---|---|---|
| (1) | (2) | (3)=(1)×(2) |
| 0 | 40 | 0 |
| 1 | 82 | 82 |
| 2 | 57 | 114 |
| 3 | 24 | 72 |
| 4 | 13 | 52 |
| 5 | 9 | 45 |
| 6 | 1 | 6 |
| মোট | 226 | 371 |

সুতরাং $\bar{x} = 371/226$ জন $= 1{\cdot}64$ জন।

এই $\bar{x}$-এর মান নির্ণয়ের একটি সরলতর পদ্ধতি পরে আলোচিত হবে।

**উদা. 4.3** 3.9 সারণীতে প্রদত্ত রাশিতথ্য থেকে মার্চ—জুলাই (1972) সময়ের জন্য কলকাতায় দৈনন্দিন সর্বোচ্চ তাপমাত্রার গড় হবে $\bar{x} = (83{\cdot}45 \times 1 + 86{\cdot}45 \times 7 + \cdots + 107{\cdot}45 \times 3)/153$ ডিঃ ফাঃ $= 95{\cdot}8028$ ডিঃ ফাঃ।

### 4.3.2 গাণিতিক গড়ের বিভিন্ন ধর্ম :

(i) যদি চলের প্রদত্ত প্রতিটি মান একটি ধ্রুবকের সমান হয় তবে চলটির গাণিতিক গড়ের মানও ঐ ধ্রুবকটির সমান হবে।

**প্রমাণ :** $x_i = a$ ( ধ্রুবক ), $i = 1(1)n$.

সুতরাং, $$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{n} a = \frac{na}{n} = a.$$

(ii) গাণিতিক গড় থেকে প্রদত্ত মানগুলির বিচ্যুতির সমষ্টির পরিমাণ শূন্য।

**প্রমাণ ঃ** [ (4.2) সূত্রটিকে গাণিতিক গড়ের সাধারণ সূত্র বলা চলে। $f_i=1,\ i=1(1)\ k,$ হলে সূত্রটি (4.1)-এ পর্যবসিত হবে। সুতরাং (4.1) কে (4.2)-এর একটি বিশেষ রূপ হিসেবে মনে করা যায়। ]

এখানে, $$\sum_{i=1}^{k} f_i(x_i-\bar{x})=\sum_{i=1}^{k} f_i x_i-\bar{x}\sum_{i=1}^{k} f_i=n\bar{x}-n\bar{x}=0$$

( প্রমাণিত )।

(iii) চলের রৈখিক রূপান্তর (linear transformation) সাধন করা হলে রূপান্তরিত চলের গড় মূল চলের গড়ের সঙ্গে অনুরূপভাবে সম্বন্ধযুক্ত হয়, অর্থাৎ, $Y=a+bX$ হলে

$$\bar{y}=a+b\bar{x} \text{ হয়।} \qquad \cdots \qquad (4.3)$$

**প্রমাণ ঃ** এখানে $y_i=a+bx_i,\ i=1(1)n.$ সুতরাং

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{k} f_i y_i=\frac{1}{n}\sum_{i=1}^{k} f_i(a+bx_i)$$

$$=\frac{a}{n}\sum_{i=1}^{k} f_i+b\frac{1}{n}\sum_{i=1}^{k} f_i x_i=a+b\bar{x} \qquad \text{( প্রমাণিত )।}$$

$$Y=a+bX \text{ অর্থাৎ, } Y=\frac{X-c}{d}\left(a=-\frac{c}{d},\ b=\frac{1}{d} \text{ লিখে}\right)$$

—এই ধরনের রূপান্তরকে মাপনার মূলবিন্দু (origin) এবং মাত্রার (scale) পরিবর্তন সাধন বলা হয়। এখানে মূলবিন্দু 0 থেকে $c$-তে এবং মাত্রা 1 থেকে $d$-তে পরিবর্তিত হয়েছে।

গাণিতিক গড়ের এই ধর্মটি অবিন্যস্ত রাশিতথ্য অথবা সমদৈর্ঘ্য শ্রেণীবিন্যাসযুক্ত পরিসংখ্যা-বিভাজন থেকে অপেক্ষাকৃত অল্প শ্রমসাপেক্ষে গাণিতিক গড় নির্ণয়ে কি-ভাবে ব্যবহার করা হয়, তা নীচের উদাহরণ দুটিতে লক্ষ্য কর।

**উদা. 4.4** হাসপাতালে নবজাতক 7টি শিশুর ওজন যথাক্রমে 3,125, 3,250, 2,960, 3,055, 3,200, 3,125 এবং 2,775 গ্রাম। এদের গড় ওজন আমরা নিম্নলিখিত পদ্ধতিতে পেতে পারি।

## সারণী 4.3

### 7 জন নবজাতকের ওজনের গাণিতিক গড় নির্ণয়

| নবজাতকের ক্রমিক সংখ্যা | ওজন ( গ্রামে ) $X$ | $Y=X-3{,}000$ |
|---|---|---|
| 1 | 3,125 | 125 |
| 2 | 3,250 | 250 |
| 3 | 2,960 | − 40 |
| 4 | 3,055 | 55 |
| 5 | 3,200 | 200 |
| 6 | 3,125 | 125 |
| 7 | 2,775 | − 225 |
| মোট | – | 490 |

এখানে $\bar{y}=490/7$ গ্রাম $=70$ গ্রাম।

আবার $y_i=x_i-3{,}000$ গ্রাম

$\therefore \quad \bar{y}=\bar{x}-3{,}000$ গ্রাম $\Rightarrow \bar{x}=3{,}000+70$ গ্রাম $=3{,}070$ গ্রাম।

এখানে 3,000 এই মানটিকে [ সাধারণভাবে $Y=a+bX$ এই রূপান্তরে $a$-কে ] যথেচ্ছ-গৃহীত মূলবিন্দু (arbitrary origin) বলে। এই বিন্দুটি গৃহীত মানগুলির যত মাঝামাঝি নেওয়া হবে, গড় নির্ণয়ে পরিশ্রমের তত লাঘব হবে। তবে এই বিন্দুটি ইচ্ছামত নির্বাচিত হলেও $\bar{x}$-এর নির্ণীত মানে কোন হেরফের হয় না।

সমদৈর্ঘ্য শ্রেণীবিন্যাসযুক্ত পরিসংখ্যা-বিভাজনের ক্ষেত্রে মূলবিন্দু ছাড়াও মাত্রা পরিবর্তনের সাহায্যে আরও কিছুটা শ্রম সঙ্কোচ করা চলে। নীচের উদাহারণটি দেখ।

**উদা 4.5** 4.3 উদাহরণে প্রদত্ত পরিসংখ্যা-বিভাজনের গড় নির্ণয় করার জন্য নিম্নলিখিত ছকে অঙ্কপাতন করা যাক।

## সারণী 4.4

### কলকাতায় মার্চ—জুলাই (1972) মাসে দৈনন্দিন সর্বোচ্চ তাপমাত্রার গড় নির্ণয়

| তাপমাত্রা ( ডিঃ ফাঃ ) | পরিসংখ্যা $f_i$ | শ্রেণী-মধ্যক $x_i$ | $y_i=\frac{x_i-95\cdot45}{3}$ | $y_if_i$ |
|---|---|---|---|---|
| 82·0 – 84·9 | 1 | 83·45 | −4 | −4 |
| 85·0 – 87·9 | 7 | 86·45 | −3 | −21 |
| 88·0 – 90·9 | 19 | 89·45 | −2 | −38 |
| 91·0 – 93·9 | 31 | 92·45 | −1 | −31 |
| 94·0 – 96·9 | 37 | 95·45 | 0 | 0 |
| 97·0 – 99·9 | 26 | 98·45 | 1 | 26 |
| 100·0 – 102·9 | 14 | 101·45 | 2 | 28 |
| 103·0 – 105·9 | 14 | 104·45 | 3 | 42 |
| 106·0 – 108·9 | 4 | 107·45 | 4 | 16 |
| মোট | 153 | – | – | 18 |

এখানে $\bar{y}=\frac{1}{153}\times 18$ ডিঃ ফাঃ $=0\cdot1176$ ডিঃ ফাঃ।

সুতরাং, $\bar{x}=95\cdot45+3\times0\cdot1176$ ডিঃ ফাঃ $=95\cdot8028$ ডিঃ ফাঃ।

এখানেও মাঝামাঝি কোন শ্রেণী-মধ্যককে পরিবর্তিত মূলবিন্দু $\left(\text{সাধারণভাবে } y=\frac{x-a}{b}\text{-তে এই রূপান্তরে } a\right)$ হিসাবে নেওয়া হয় এবং পরিবর্তিত মাপনামাত্রা ($b$) হিসাবে নেওয়া হয় সাধারণ শ্রেণীদৈর্ঘ্যটিকে। শ্রেণীগুলি সমদৈর্ঘ্য না হলেও দৈর্ঘ্যগুলির গ. সা. গু.-কে পরিবর্তিত মাপনামাত্রা হিসাবে নিয়ে কিছুটা শ্রমসঙ্কোচ করা যায়।

(iv) $n_1$ ও $n_2$ সংখ্যক মানসম্পন্ন দুটি গোষ্ঠীর গাণিতিক গড় যথাক্রমে $\bar{x}_1$

এবং $\bar{x}_2$ হলে, এই $n_1+n_2$ সংখ্যক মানের সার্বিক গড় (grand mean) হবে

$$\bar{x}=\frac{n_1\bar{x}_1+n_2\bar{x}_2}{n_1+n_2}. \quad \cdots \quad (4.4)$$

**প্রমাণ।** মনে কর প্রথম গোষ্ঠীর মানগুলি

$x_{11}, x_{12}, \cdots\cdots x_{1n_1}$ ও তাদের গড় $=\bar{x}_1$,

দ্বিতীয় গোষ্ঠীর মানগুলি

$x_{21}, x_{22}, \cdots\cdots, x_{2n_2}$ ও তাদের গড় $=\bar{x}_2$,

এবং এই $(n_1+n_2)$ সংখ্যক রাশিগুচ্ছ একত্রিত করলে, তাদের সার্বিক গড় $=\bar{x}$.

তাহলে, $$\bar{x}_i=\frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}, \quad i=1, 2.$$

এবং, $$(n_1+n_2)\bar{x}=\sum_{i=1}^{n_1} x_{1j}+\sum_{j=1}^{n_2} x_{2j}$$
$$=n_1\bar{x}_1+n_2\bar{x}_2.$$ (প্রমাণিত)।

(4.4) সূত্রটি সরাসরি দুই-এর বেশী গোষ্ঠীর ক্ষেত্রে সম্প্রসারিত করা চলে। মনে কর, $k$ সংখ্যক গোষ্ঠীর $i$-তমটির মানসংখ্যা $n_i$ এবং গড় $\bar{x}_i$, $i=1\ (1)\ k$.

অতএব $\sum_{i=1} n_i$ সংখ্যক মানের সার্বিক গড়

$$\bar{x}=\sum_{i=1}^{k} n_i\bar{x}_i \Big/ \sum_{i=1} n_i. \quad \cdots \quad (4.5)$$

(v) একটি চল অন্য একাধিক চলের সঙ্গে রৈখিকভাবে (linearly) যুক্ত হলে, প্রথমোক্ত চলের গাণিতিক গড়ও শেষোক্ত চলগুলির গাণিতিক গড়গুলির সঙ্গে অনুরূপভাবে সম্বন্ধযুক্ত হবে।

অর্থাৎ, $x_i=a+by_i+cz_i+\cdots+hw_i$ হলে

$\bar{x}=a+b\bar{y}+c\bar{z}+\cdots+h\bar{w}$ হবে। $\cdots$ (4.6)

**প্রমাণ।** এখানে $$\bar{x}=\frac{1}{n}\sum_{i=1}^{k} f_i\left(a+by_i+cz_i+\cdots+hw_i\right)$$
$$=a+b\bar{y}+c\bar{z}+\cdots+h\bar{w}.$$ (প্রমাণিত)।

## 4.4 ভগ্নাংশক (quantile বা fractile) এবং মধ্যমা (median) ঃ

**4.4.1 সংজ্ঞা ঃ** চলের প্রদত্ত মানগুলি ঊর্ধ্বগ বা নিম্নগ ক্রমানুসারে সাজানো হলে যে মানটি বিভাজনটিকে $p:(1-p)$ অনুপাতে ভাগ করে সেটিকে চলের $p$-তম **ভগ্নাংশক** বলা হয়। $p=\cdot 5$ হলে সংশ্লিষ্ট ভগ্নাংশকটিকে **মধ্যমা** বলে। মধ্যমা একটি বহুল ব্যবহৃত মধ্যগামিতা-মাপক, তাই বর্তমান অনুচ্ছেদে এই বিশেষ ভগ্নাংশকটি সম্বন্ধে বিস্তারিত আলোচনা করা হবে।

$p=\cdot 25$ এবং $\cdot 75$ হলে ভগ্নাংশকগুলি যথাক্রমে প্রথম ও তৃতীয় **চতুর্থক** (quartile) নামে পরিচিত। স্পষ্টতঃই, দ্বিতীয় চতুর্থক হচ্ছে মধ্যমা। মধ্যমা চলের প্রদত্ত বিভাজনটিকে সমদ্বিখণ্ডিত করে। তেমনি চতুর্থকগুলি একযোগে বিভাজনটিকে করে সমচতুষ্খণ্ডিত। অনুরূপভাবে **দশমক** (decile) এবং **শততমক** (percentile)-এর সংজ্ঞা দেওয়া যায়। লক্ষণীয়, প্রতিটি ভগ্নাংশকই এক একটি বিবরণাত্মক মাপক।

সংজ্ঞানুযায়ী ক্রমানুসারে সাজানো মানগুলির ঠিক মধ্যবর্তীটিই মধ্যমা, কারণ এই মানটিই বিভাজনটিকে সমদ্বিখণ্ডিত করে।

**4.4.2 মধ্যমা-নির্ণয় ঃ** বিভিন্ন পরিস্থিতিতে কি-ভাবে মধ্যমা নির্ণয় করা হয়, দেখা যাক।

প্রথমে মনে কর, চলের $x_1, x_2, \ldots, x_n$ এই $n$টি অবিন্যস্ত মান দেওয়া আছে। এগুলি ঊর্ধ্বগ ক্রমানুসারে সাজিয়ে লেখা হ'ল $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ এইভাবে। এখানে $x_{(1)}$ হচ্ছে $x_1, x_2, \ldots, x_n$ এদের মধ্যে ক্ষুদ্রতম, $x_{(2)}$ পরবর্তী ক্ষুদ্রতম, $\cdots$ এবং $x_{(n)}$ এদের মধ্যে বৃহত্তম। এখন $n=2m+1$, অর্থাৎ অযুগ্ম হলে, স্পষ্টতঃই মধ্যমা $\tilde{x}=x_{(m+1)}$. আর $n=2m$, অর্থাৎ যুগ্ম হলে, মধ্যবর্তী মান পাওয়া যাবে দুটি—$x_{(m)}$ এবং $x_{(m+1)}$. প্রকৃতপক্ষে এই দুটি মানের মধ্যবর্তী যে কোন মানকেই মধ্যমা বলা চলে এক্ষেত্রে। সাধারণতঃ মান-দুটির গাণিতিক গড়কেই মধ্যমা হিসাবে নেওয়াই প্রথা, অর্থাৎ, $\tilde{x}=\frac{1}{2}[x_{(m)}+x_{(m+1)}]$. অবশ্য বিচ্ছিন্ন চলের ক্ষেত্রে যেখানে ভগ্নাংশবিশিষ্ট মান অর্থহীন, সেখানে অনেক সময় এই দুটি মানকেই [ অর্থাৎ $x_{(m)}$ ও $x_{(m+1)}$-কে ] মধ্যমা হিসাবে ধরা হয়।

**উদা. 4.6** 4.1 সারণীতে প্রদত্ত মানগুলিকে উর্ধ্বগ ক্রমানুসারে সাজালে দাঁড়ায়

7·2, 7·6, 8·2, 8·3, 8·4, 8·5, 8·6, 9·0, 9·1, 9·5.

এখানে $n=10$ ( যুগ্মসংখ্যা )।

অতএব $\tilde{x}=\frac{1}{2}[x_{(5)}+x_{(6)}]$

$=\frac{1}{2}[8·4+8·5]$ কুইণ্টাল $=8·45$ কুইণ্টাল।

প্রদত্ত দশখণ্ড জমির সঙ্গে আর এক খণ্ড, যার একরপ্রতি ফলন 7·9 কুইণ্টাল নেওয়া হলে, ক্রমানুসারে সাজানো মানগুলি দাঁড়াবে :

7·2, 7·6, 7·9, 8·2, 8·3, 8·4, 8·5, 8·6, 9·0, 9·1, 9·5.

এখানে $n=11$ ( বিযুগ্মসংখ্যা )। সুতরাং মধ্যমা $\tilde{x}=x_{(6)}=8·4$ কুইণ্টাল।

প্রদত্ত রাশিতথ্য যদি শ্রেণীবিন্যস্ত আকারে থাকে এবং এক-একটি মান সূচিত করে এক-একটি শ্রেণী, তাহলে সহজেই ক্রমযৌগিক পরিসংখ্যা সারণী থেকে মধ্যমা চিহ্নিত করা যায়। এক্ষেত্রে মানগুলি সাধারণতঃ উর্ধ্বগ বা নিম্নগ ক্রমানুসারে সাজানোই থাকে; ক্রমযৌগিক পরিসংখ্যার বিচারে তাই সহজেই এগুলিকে ক্রমিক সংখ্যার সাহায্যে চিহ্নিত করা যায়। সুতরাং বিশেষ ক্রমিকসংখ্যা-সম্পন্ন মানটি খুঁজে নেওয়া অনায়াসেই সম্ভব হয়।

**উদা. 4.7** 3.4 সারণীতে প্রদত্ত দিনপ্রতি অনুপস্থিত শিক্ষকসংখ্যার মধ্যমা 3.5 সারণী থেকে সহজেই পাওয়া যায়। এখানে মানগুলি উর্ধ্বগ ক্রমানুসারে সাজানো আছে। 0, 1, 2,···, 6 এই মানগুলির ক্রমযৌগিক পরিসংখ্যা যথাক্রমে 40, 122, 179,···, 226—অর্থাৎ 1 থেকে 40-তম মানগুলির প্রত্যেকটি 0, 41-তম থেকে 122-তম মানগুলির প্রত্যেকটি 1, ··· ইত্যাদি। এখানে $n=226$; সুতরাং 113-তম ও 114-তম মানের গড়ই মধ্যমা। স্পষ্টতঃই 113-তম ও 114-তম উভয় মানই 1, সুতরাং মধ্যমা $\tilde{x}=1$.

এক্ষেত্রে একই নিয়মে অন্যান্য ভগ্নাংশকও পাওয়া যেতে পারে। বস্তুতঃ $p$-তম ভগ্নাংশক $z_p$-এর সূত্র হবে $z_p=x_{(\overline{n+1}p)}$. ··· (4.7)

এখানে $(n+1)p$ যদি অখণ্ড সংখ্যা হয় তাহলে $p$-তম ভগ্নাংশক হবে উর্ধ্ব ক্রমানুসারে সাজানো $(n+1)p$-তম মানটি। যদি অখণ্ড সংখ্যা না হয়, মনে কর, $(n+1)p=k+h$, যেখানে $k$ একটি অখণ্ড সংখ্যা এবং $0<h<1$. এক্ষেত্রে আলোচ্য ভগ্নাংশকটি হবে সেই বিন্দুটি যা $k$-তম এবং $(k+1)$-তম মানের মধ্যবর্তী অন্তরটিকে $h:1-h$ অনুপাতে ভাগ করে।

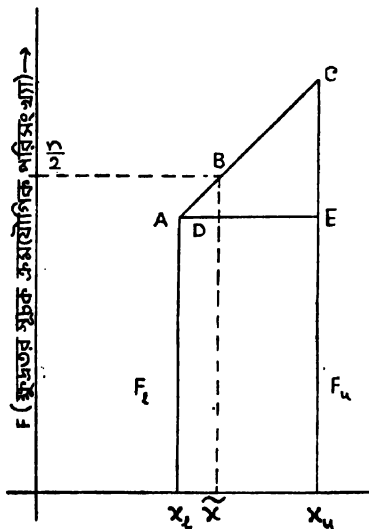
$n$-এর মান খুব কম হলে সাধারণতঃ মধ্যমা ও চতুর্থক ছাড়া অন্যান্য ভগ্নাংশক নির্ণয় করা হয় না।

আলোচ্য উদাহরণে $Q_1 = z_{.25} = x_{(227\times.25)} = x_{(56.75)}$ অর্থাৎ, $x_{(56)}$ ও $x_{(57)}$-এর মধ্যে যে বিন্দুটি এই তুটি মানের মধ্যবর্তী অন্তরকে 75 : 25 অনুপাতে ভাগ করে।

কিন্তু $x_{(56)} = x_{(57)} = 1$. সুতরাং $Q_1 = 1$.

অনুরূপভাবে $Q_3 = z_{.75} = 2$.

আলোচ্য চলটি অবিচ্ছিন্ন হলে সাধারণতঃ এক-একটি শ্রেণী গঠিত হয় একাধিক মান নিয়ে। এক্ষেত্রে $n/2$-তম মানটি সঠিকভাবে চিহ্নিত করা সম্ভব নয়, একথা আগেই বলা হয়েছে। সুতরাং মধ্যমারও যথার্থ মান পাওয়া যায় না। অবশ্য রৈখিক অন্তঃপ্রক্ষেপণ পদ্ধতির (linear interpolation) সাহায্যে মধ্যমার একটি আসন্ন মান পাওয়া যায়।

এক্ষেত্রে ক্রমযৌগিক পরিসংখ্যার বিচারে প্রথমে মধ্যমাশ্রেণীটি (median-class) চিহ্নিত করা হয়। নিম্নতম যে শ্রেণীটির ক্রমযৌগিক পরিসংখ্যা ( ক্ষুদ্রতর-সূচক ) $n/2$ অপেক্ষা বড়, সেটিই মধ্যমাশ্রেণী। মনে কর $x_l$ এবং $x_u$ যথাক্রমে

চিত্র 4.1

শ্রেণীবিন্যস্ত পরিসংখ্যা-বিভাজন থেকে মধ্যমা নির্ণয়।

মধ্যমাশ্রেণীর অধঃ- ও ঊর্ধ্বসীমান্ত এবং $F_l$ ও $F_u$ সংশ্লিষ্ট ক্রমযৌগিক পরিসংখ্যা। এখানে লক্ষণীয়, $x_l$ মধ্যমাশ্রেণীর পূর্ববর্তী শ্রেণীটির ঊর্ধ্বসীমান্তও বটে, সুতরাং $F_l$ পূর্ববর্তী শ্রেণীর ক্ষুদ্রতর-সূচক ক্রমযৌগিক পরিসংখ্যা। স্পষ্টতঃই, $F_l < n/2 < F_u$. এখন মধ্যমাশ্রেণীর অন্তর্গত মানগুলি শ্রেণী-অন্তরটিতে সমভাবে নিবেশিত এই স্বীকরণসাপেক্ষে $x_l$ এবং $x_u$-এর মধ্যে ক্রমযৌগিক পরিসংখ্যা রেখাটিকে সরলরেখা ভাবা যেতে পারে। সংজ্ঞানুসারে 4.1 চিত্রে যে বিন্দুটির কোটি (ordinate) $n/2$ সেটির ভুজই (abscissa) মধ্যমা। স্পষ্টতঃই $\triangle ABD$ ও $\triangle ACE$ সদৃশ। সুতরাং

$$\frac{AD}{AE}=\frac{BD}{CE}$$

অর্থাৎ, $$\frac{\tilde{x}-x_l}{x_u-x_l}=\frac{n/2-F_l}{F_u-F_l}$$

অর্থাৎ, $$\tilde{x}=x_l+\frac{n/2-F_l}{f_0}\times c, \qquad \cdots \quad (4.8)$$

যেখানে $c$ ও $f_0$ মধ্যমাশ্রেণীর যথাক্রমে প্রসার ও পরিসংখ্যা।

**উদা. 4.8** 3.10 সারণীতে প্রদত্ত রাশিতথ্যের ক্ষেত্রে প্রথমে ক্রমযৌগিক পরিসংখ্যা সারণী থেকে 93·95 – 96·95 এই শ্রেণীটিকে মধ্যমাশ্রেণী হিসাবে চিহ্নিত করা হ'ল, কেননা এই শ্রেণীতেই $\frac{153}{2}$-তম মানটি অন্তর্ভুক্ত। এরপর (4.8) সূত্রটি ব্যবহার ক'রে পাওয়া যায়

$$\tilde{x}=93{\cdot}95+\frac{76{\cdot}5-58}{37}\times 3$$

$$=93{\cdot}95+\frac{18{\cdot}5}{37}\times 3=93{\cdot}95+1{\cdot}5000=95{\cdot}4500 \text{ (ডিঃ ফাঃ)।}$$

একই পদ্ধতিতে যে কোন ভগ্নাংশকের মান নির্ণয় করা যায়। $p$-তম ভগ্নাংশক $z_p$-এর সূত্র

$$z_p=x_l+\frac{np-F_l}{f_0}\times c, \qquad (4.9)$$

যেখানে, $x_l$ হচ্ছে $p$-তম ভগ্নাংশক-শ্রেণীর অধঃসীমান্ত, $F_l$ হচ্ছে $x_l$-এর ক্রমযৌগিক পরিসংখ্যা এবং $f_0$ ও $c$ যথাক্রমে শ্রেণীটির পরিসংখ্যা এবং দৈর্ঘ্য।

**উদা. 4.9** 3.10 সারণীতে প্রদত্ত রাশিতথ্যের ক্ষেত্রে প্রথম ও তৃতীয় চতুর্থকের মান পাওয়া যায়

$$Q_1 = 90{\cdot}95 + \frac{38{\cdot}25 - 27}{31} \times 3$$

$$= 90{\cdot}95 + 1{\cdot}0887 = 92{\cdot}0387 \text{ ( ডিঃ ফাঃ )।}$$

এবং

$$Q_3 = 96{\cdot}95 + \frac{114{\cdot}75 - 95}{26} \times 3$$

$$= 96{\cdot}95 + 2{\cdot}2788 = 99{\cdot}2288 \text{ ( ডিঃ ফাঃ )।}$$

### 4.4.3 লৈখিক পদ্ধতিতে ভগ্নাংশক ও মধ্যমা নির্ণয় :

পরিসংখ্যা-বিভাজনের ক্রমযৌগিক পরিসংখ্যা-রেখা ( ক্ষুদ্রতর-সূচক অথবা বৃহত্তর-সূচক ) থেকে সহজেই মধ্যমা এবং অন্যান্য ভগ্নাংশকের মান নির্ণয় করা সম্ভব। ক্রমযৌগিক পরিসংখ্যা-রেখা এবং অনুভূমিক সরলরেখা $Y = np$ এই দুটির ছেদবিন্দুর ভুজ স্পষ্টতঃই চলটির $p$-তম ভগ্নাংশক। সুতরাং $Y = n/2$ এবং ক্রমযৌগিক পরিসংখ্যা-রেখার ছেদবিন্দুর ভুজই মধ্যমা।

3.7 চিত্রে 3.8 সারণীতে প্রদত্ত রাশিতথ্যের মধ্যমা, $Q_1$ এবং $Q_3$ লৈখিক পদ্ধতিতে নির্ণয় করা হয়েছে। লক্ষ্য কর, এই পদ্ধতিতে নির্ণীত মানগুলি ( যথাক্রমে, 95·45, 91·55 এবং 99·35 ডিঃ ফাঃ ) পূর্বে নির্ণীত মানগুলির খুব কাছাকাছি।

একই চিত্রে ক্ষুদ্রতর-সূচক এবং বৃহত্তর-সূচক পরিসংখ্যা-রেখা অঙ্কিত হলে রেখা-দুটির ছেদবিন্দুর কোটি স্পষ্টতঃই $n/2$—সুতরাং বিন্দুটির ভুজই মধ্যমা মান। 3.7 চিত্রে লক্ষ্য কর, ক্রমযৌগিক পরিসংখ্যা-রেখা দুটির ছেদবিন্দুর ভুজ 95·45।

### 4.4.4 মধ্যমার একটি বিশেষ ধর্ম :

কোন চলের যদি এমনভাবে রূপান্তর সাধন করা হয় যে, মূল চলের প্রদত্ত মানগুলির ক্রমটি (order) রূপান্তরিত চলের ক্ষেত্রেও অক্ষুণ্ণ থাকে, তাহলে স্পষ্টতঃই রূপান্তরিত চলের মধ্যমাটিও হবে মূল চলের মধ্যমার অনুরূপ রূপান্তর। মনে কর, $X$-এর প্রদত্ত মানগুলি যথাক্রমে 4, 6, 8, 9, 11 ; সুতরাং $\tilde{x} = 8$. এখন $Y = X^2$ হলে রূপান্তরিত চলের মানগুলি দাঁড়াবে 16, 36, 64, 81 এবং 121. স্পষ্টতঃই $\tilde{y} = 64 = \tilde{x}^2$.

লক্ষ্য কর, $x$-এর প্রদত্ত মানগুলির কিছু ধনাত্মক, কিছু ঋণাত্মক হলে $Y = X^2$ এই রূপান্তরে মূল চলের মানক্রমটি রূপান্তরিত চলের ক্ষেত্রে অক্ষুণ্ণ থাকে না।

## 4.5 ভূয়িষ্ঠক বা সংখ্যাগরিষ্ঠ মান (mode) :

সংগৃহীত মানগুলির কেন্দ্রীভবনের প্রবণতাকে চলের মধ্যগামিতা আখ্যা দেওয়া হয়েছে। সুতরাং যে বিন্দুটিতে কেন্দ্রীভবন সর্বাপেক্ষা বেশী সেটিকে স্বভাবতঃই মধ্যগামিতার একটি মাপক হিসাবে ব্যবহার করার কথা ভাবা যেতে পারে। এই ধারণা থেকেই মধ্যগামিতা-মাপক হিসাবে ভূয়িষ্ঠক বা সংখ্যাগরিষ্ঠ মানের প্রচলন হয়েছে।

বিচ্ছিন্ন চলের ক্ষেত্রে চলের যে মানটির পরিসংখ্যা সর্বাধিক, সেটিকেই বলা হয় ভূয়িষ্ঠক। স্পষ্টতঃই, স্বল্পসংখ্যক কয়েকটি মান দেওয়া থাকলে ভূয়িষ্ঠক নির্ণয় করা সম্ভব নাও হতে পারে এবং তা উচিতও নয়। বিচ্ছিন্ন চলের পরিসংখ্যা-বিভাজনে একাধিক মান সর্বোচ্চ পরিসংখ্যা-সম্পন্ন হলে ভূয়িষ্ঠকের অস্তিত্ব নাই ধ'রে নেওয়া হয়।

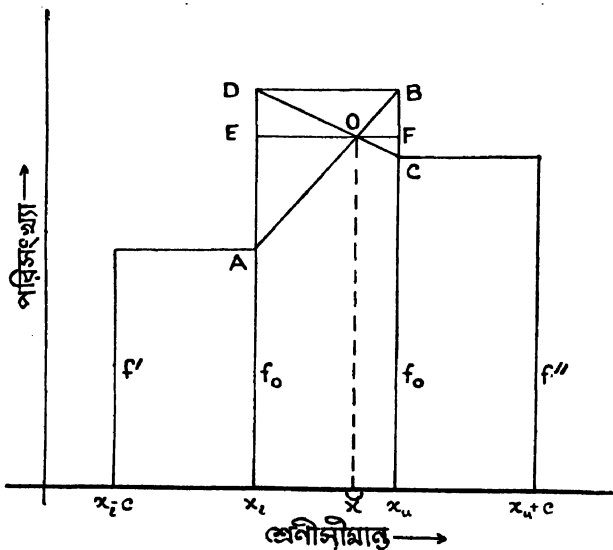
**উদা. 4.10** 3.4 সারণীতে দেখা যাচ্ছে $X = 1$ এই মানটির পরিসংখ্যা সর্বোচ্চ (82) ; সুতরাং এক্ষেত্রে ভূয়িষ্ঠক $= 1$.

অবিচ্ছিন্ন চলের ক্ষেত্রে একটি একক (single) মানের সর্বোচ্চ পরিসংখ্যা-সম্পন্ন হওয়া সম্ভব নয় বোধগম্য কারণেই। সুতরাং অবিচ্ছিন্ন চলের ক্ষেত্রে ভূয়িষ্ঠকের এই সংজ্ঞাটি প্রযোজ্য নয়। এক্ষেত্রে পরিসংখ্যা-রেখাটি আঁকা সম্ভব হলে $X$-এর যে মানের জন্য রেখাটি সর্বোচ্চ কোটিবিশিষ্ট, অর্থাৎ রেখাটির চরমাবস্থা (maximum)—সেই বিন্দুতেই কেন্দ্রীভবনের মাত্রা সর্বাধিক, তাই এটিকেই বলা হয় ভূয়িষ্ঠক।

কোন কোন চলের পরিসংখ্যা-রেখার দুই বা ততোধিক স্থানীয় চরমাবস্থা (local maxima) থাকা সম্ভব। সেক্ষেত্রে চলটিকে দ্বিভূয়িষ্ঠক (bimodal) বা বহুভূয়িষ্ঠক (multimodal) বলা হবে, যদি যথাক্রমে দুই বা ততোধিক স্থানীয় চরমাবস্থা থাকে। চলের প্রদত্ত পরিসংখ্যা-বিভাজন দ্বিভূয়িষ্ঠক বা বহুভূয়িষ্ঠক হলে এমন সন্দেহ হওয়া স্বাভাবিক যে বিভাজনটি এমন দুই বা ততোধিক গোষ্ঠী-সংক্রান্ত তথ্যের সংমিশ্রণে উদ্ভূত হয়েছে, যাদের মধ্যগামিতা লক্ষণীয়ভাবে ভিন্ন। যেমন বেশ কিছুসংখ্যক ভারতীয় পুরুষ ও নারী একত্রিত ক'রে তাদের উচ্চতার যে পরিসংখ্যা-বিভাজন পাওয়া যাবে, সেটির দ্বিভূয়িষ্ঠক হওয়ার সম্ভাবনা খুব বেশী।

অন্তর্বিষম রাশিতথ্যের (heterogeneous data) ক্ষেত্রে এই ধরনের পরিস্থিতির উদ্ভব ঘটে।

এখন মোট পরিসংখ্যার পরিমাণ সীমাহীনভাবে বৃহৎ হলে তবেই পরিসংখ্যা-রেখাটি অঙ্কন করা সম্ভব। কিন্তু বাস্তবক্ষেত্রে সাধারণতঃ চলের সীমিতসংখ্যক মান দেওয়া থাকে। সুতরাং প্রশ্ন : সেক্ষেত্রে ভূয়িষ্ঠক কি-ভাবে নির্ণয় করা হবে? শ্রেণীবিন্যস্ত পরিসংখ্যা-বিভাজনে অবশ্য সহজেই সর্বোচ্চ পরিসংখ্যা-সম্পন্ন শ্রেণীটিকে ভূয়িষ্ঠক-শ্রেণী (modal class) হিসাবে চিহ্নিত করা যায়। কিন্তু সাধারণতঃ একটি শ্রেণী-অন্তরের পরিবর্তে ভূয়িষ্ঠকের একটি একক মানেরই বেশী প্রয়োজন হয়। আগেই বলা হয়েছে এক্ষেত্রে ভূয়িষ্ঠকের যথার্থ মান পাওয়া সম্ভব নয়। আসন্ন মান হিসাবে ভূয়িষ্ঠক-শ্রেণীর মধ্যকটি গ্রহণ করা যেতে পারে। যেমন 3.9 সারণীতে প্রদত্ত পরিসংখ্যা-বিভাজনের ক্ষেত্রে 95·45 ডিঃ ফাঃ।

এখন সমদৈর্ঘ্য শ্রেণীবিন্যাসে $x_l$ ও $x_u$ সীমান্তবিশিষ্ট ভূয়িষ্ঠক-শ্রেণীর পূর্ববর্তী এবং পরবর্তী শ্রেণী-পরিসংখ্যা, ধরা যাক, যথাক্রমে $f'$ এবং $f''$, পরস্পর সমান হলে ভূয়িষ্ঠক হিসাবে $\frac{1}{2}(x_l+x_u)$ নেওয়া যুক্তিযুক্ত হবে। অন্যথায়, পরিসংখ্যা-বিভাজনের আয়তলেখটি লক্ষ্য করলে দেখা যাবে, $f'$ অপেক্ষা $f''$ বড় ( ছোট )

চিত্র 4.2

শ্রেণীবিন্যস্ত পরিসংখ্যা-বিভাজন থেকে ভূয়িষ্ঠক নির্ণয়।

হলে লব্ধ পরিসংখ্যা-রেখার শীর্ষ-নির্দেশক মানটির, অর্থাৎ ভূয়িষ্ঠকের, $x_u$ ($x_l$)-এর দিকে সরে যাওয়ার প্রবণতা রয়েছে ( চিত্র 4.2 )। প্রকৃতপক্ষে ভূয়িষ্ঠক, ধরা যাক $\breve{x}$, মোটামুটিভাবে ভূয়িষ্ঠক শ্রেণী-অন্তরটিকে $f_0 - f' : f_0 - f''$ ( $f_0$ = ভূয়িষ্ঠক-শ্রেণীর পরিসংখ্যা ) অনুপাতে বিভক্ত করে। ওপরের চিত্রটি লক্ষ্য করলে ব্যাপারটি আরও স্পষ্ট হবে।

এখানে $\triangle OED$ এবং $\triangle OFC$ এই দুটি সদৃশ ত্রিভুজ থেকে পাওয়া যায়

$$\frac{OD}{OC} = \frac{OE}{OF}.$$

এবং $\triangle ODA$ এবং $\triangle OCB$ এই দুটি সদৃশ ত্রিভুজ থেকে পাওয়া যায়

$$\frac{OD}{OC} = \frac{AD}{BC}.$$

সুতরাং, $$\frac{OE}{OF} = \frac{AD}{BC}$$

অর্থাৎ, $$\frac{\breve{x} - x_l}{x_u - \breve{x}} = \frac{f_0 - f'}{f_0 - f''}$$

অর্থাৎ, $$\breve{x} = x_l + \frac{f_0 - f'}{2f_0 - f' - f''} \times c, \qquad \cdots \qquad \cdots \qquad (4.10)$$

যেখানে $c$ ভূয়িষ্ঠক শ্রেণীর দৈর্ঘ্য।

ভূয়িষ্ঠক-সংক্রান্ত আলোচনা শেষ করার আগে একটি বিষয়ে দৃষ্টি আকর্ষণ করা প্রয়োজন। পরিসংখ্যা-বিভাজনে প্রদত্ত রাশিতথ্যে বিশেষ একটি শ্রেণীর পরিসংখ্যা অন্যান্য শ্রেণীগুলির পরিসংখ্যা থেকে যথেষ্ট পরিমাণে বেশী হলে তবেই সংশ্লিষ্ট শ্রেণীটিকে সন্দেহাতীতভাবে ভূয়িষ্ঠক-শ্রেণী হিসাবে চিহ্নিত করা চলে। কিন্তু পার্থক্যের পরিমাণ খুব কম হলে, বিশেষ ক'রে অবিচ্ছিন্ন চলের ক্ষেত্রে যেহেতু শ্রেণীবিন্যাস অনেকটাই কৃত্রিম এবং ব্যক্তিনির্ভর, এমন সন্দেহ হওয়া খুবই স্বাভাবিক যে শ্রেণীগুলি একটু অন্যভাবে নেওয়া হলে সর্বোচ্চ পরিসংখ্যা দাঁড়াত হয়ত নিকটবর্তী অন্য একটি শ্রেণীর। সুতরাং এই ধরনের পরিস্থিতিতে ভূয়িষ্ঠক-নির্ণয়ে কিছুটা সতর্কতা অবলম্বন করা প্রয়োজন।

**উদা. 4.10** 3.9 সারণীর রাশিতথ্যের ক্ষেত্রে ভূয়িষ্ঠকের মান নির্ণয় করা যাক। এক্ষেত্রে $x_l = 93{\cdot}95, f_0 = 37, f' = 31, f'' = 26$ এবং $c = 3$.

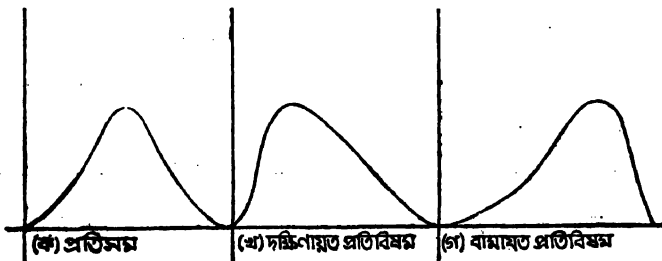
সুতরাং, $$\breve{x} = 93{\cdot}95 + \frac{3(37-31)}{2 \times 37 - 31 - 26} \text{ ডিঃ ফাঃ}$$

$$= 95{\cdot}0088 \text{ ডিঃ ফাঃ।}$$

## 4.6 গাণিতিক গড়, মধ্যমা এবং ভূয়িষ্ঠকের মধ্যে অবেক্ষণভিত্তিক সম্পর্ক :

কোন অবিচ্ছিন্ন চলের ওপর সংগৃহীত রাশিতথ্যের আয়তলেখটি এবং আয়তলেখের ওপর পরিসংখ্যা-রেখাটি অঙ্কন করা যাক। কল্পনা কর, আয়তলেখটি তীক্ষ্ণধার ধাতব পাতের ওপর দণ্ডায়মান, বিভিন্ন শ্রেণীগুলির পরিসংখ্যা-নির্দেশী প্রতিটি আয়তক্ষেত্রও অনুরূপ ধাতুতে নির্মিত। সংজ্ঞানুসারে, ভূমির ওপর সমগ্র আয়তলেখটির ভরকেন্দ্রই রাশিতথ্যের গড়, সমগ্র বিভাজনটিকে সমদ্বিখণ্ডনকারী বিন্দুটিই মধ্যমা এবং ভূমিগত যে বিন্দুটিতে পরিসংখ্যা-রেখা সর্বোচ্চ কোটিবিশিষ্ট, সেটিই ভূয়িষ্ঠক।

কোন বিচ্ছিন্ন চলের ক্ষেত্রে $h$-এর যে কোন গ্রাহ্য মানের জন্য যদি $x_0+h$ এবং $x_0-h$-এর পরিসংখ্যা সমান হয় তাহলে বিভাজনটিকে $x_0$-কেন্দ্রিক **প্রতিসম** (symmetrical about $x_0$) বলা হয়। অবিচ্ছিন্ন চলের ক্ষেত্রে চলের একটি একক মানের পরিসংখ্যার প্রশ্নটি অর্থহীন। এক্ষেত্রে বিভাজনটি $x_0$-কেন্দ্রিক প্রতিসম হবে যদি চলটির পরিসংখ্যা-রেখায় $h$-এর যে কোন মানের জন্য $x_0+h$ এবং $x_0-h$ বিন্দু-দুটিতে কোটিদ্বয় সমান হয়। অল্পসংখ্যক রাশিতথ্য থেকে পরিসংখ্যা-রেখা পাওয়া যায় না, আগেই বলা হয়েছে। তবে সংশ্লিষ্ট বিভাজনটির আয়তচিত্রের আকৃতি থেকে বিভাজনটি প্রতিসম কিনা মোটামুটি আন্দাজ করা যায়।

কোন বিভাজন প্রতিসম না হলে তাকে বলা হয় **প্রতিবিষম** (skew) বিভাজন। প্রতিসম বিভাজনের পরিসংখ্যা-রেখাটি **ঘণ্টাকৃতি-বিশিষ্ট** (bell-shaped)—এর পুচ্ছ-দুটি সমান দৈর্ঘ্যের এবং সমভাবে ন্যস্ত [ চিত্র 4.3 (ক) ]। স্পষ্টতঃই, প্রতিবিষম বিভাজনের পুচ্ছ-দুটির দৈর্ঘ্য অসমান—ডানদিকের

চিত্র 4.3

**(ক) প্রতিসম, (খ) দক্ষিণায়ত প্রতিবিষম ও (গ) বামায়ত প্রতিবিষম পরিসংখ্যা-রেখা।**

অথবা বামদিকের পুচ্ছটি অধিকতর বিস্তৃত হলে যথাক্রমে পাওয়া যায় **দক্ষিণায়ত** ( অথবা **ধনাত্মক** ) এবং **বামায়ত** ( বা **ঋণাত্মক** ) **প্রতিবিষম** (positively and negatively skew) বিভাজন [ যথাক্রমে চিত্র 4.3 (খ) ও (গ) ] ।

অনুচ্ছেদের শুরুতে আলোচিত গড়, মধ্যমা এবং ভূয়িষ্ঠকের প্রকৃতি থেকে সহজেই বলা যায় $x_0$-কেন্দ্রিক প্রতিসম বিভাজনের ক্ষেত্রে

$$\bar{x} = \tilde{x} = \breve{x} = x_0. \quad \cdots \quad \cdots \quad (4.11)$$

দক্ষিণায়ত এবং বামায়ত প্রতিবিষম বিভাজনের ক্ষেত্রে যথাক্রমে

$$\bar{x} > \tilde{x} > \breve{x} \quad \cdots \quad \cdots \quad (4.12a)$$

এবং

$$\bar{x} < \tilde{x} < \breve{x} \quad \cdots \quad \cdots \quad (4.12b)$$

অসমতা সম্পর্ক-দুটি যে সত্য, তা 4.3 চিত্রগুলি লক্ষ্য করলেই বোঝা যাবে। গাণিতিক গড়, মধ্যমা ও ভূয়িষ্ঠকের এই আপেক্ষিক অবস্থিতি সহজে মনে রাখা যায়, ইংরেজী অভিধানে এদের ইংরেজী প্রতিশব্দগুলির ( যথাক্রমে mean, median এবং mode) আপেক্ষিক অবস্থিতি থেকে। অভিজ্ঞতা থেকে দেখা গেছে স্বল্পপ্রতিবিষম যে কোন বিভাজনের ক্ষেত্রে

$$(\bar{x} - \breve{x}) \simeq 3(\bar{x} - \tilde{x}) \quad \cdots \quad \cdots \quad (4.13)$$

এই অবেক্ষণভিত্তিক (empirical) আসন্ন সম্পর্কটি সত্য।

পূর্ববর্তী অনুচ্ছেদের আলোচনা থেকে দেখা গেছে অনেক সময় প্রদত্ত পরিসংখ্যা-বিভাজনের ভূয়িষ্ঠক সহজে নির্ণয় করা যায় না। সেক্ষেত্রে $\bar{x}$ এবং $\tilde{x}$-এর মান জানা থাকলে (4.13) সূত্রটি ব্যবহার ক'রে $\breve{x}$-এর আসন্ন মান নির্ণয় করা যেতে পারে।

4.10 উদাহরণে 3.9 সারণীতে প্রদত্ত পরিসংখ্যা-বিভাজনের ভূয়িষ্ঠক নির্ণয় করা হয়েছে। এখন 4.13 সূত্রটি ব্যবহার ক'রে ভূয়িষ্ঠকের মান কত হয় দেখা যাক।

$$\breve{x} \simeq 3\tilde{x} - 2\bar{x}$$

$$= 3 \times 95{\cdot}4500 - 2 \times 95{\cdot}8028 \text{ ডিঃ ফাঃ}$$

$$= 94{\cdot}7444 \text{ ডিঃ ফাঃ।}$$

লক্ষ্য কর, অবেক্ষণভিত্তিক সম্পর্ক থেকে পাওয়া ভূয়িষ্ঠকের মানটি 4.10 উদাহরণে নির্ণীত মানের মোটামুটি কাছাকাছি।

## 4.7 গাণিতিক গড়, মধ্যমা এবং ভূয়িষ্ঠকের মধ্যে তুলনা :

একটি আদর্শ মধ্যগামিতা-মাপকের সংজ্ঞা সুস্পষ্ট এবং দ্ব্যর্থহীনভাবে নির্দিষ্ট হওয়া উচিত এবং কোন প্রদত্ত পরিস্থিতিতে এটির মানও সুনির্দিষ্ট হওয়া উচিত। আলোচ্য তিনটি মাপকের সংজ্ঞা সুস্পষ্ট হলেও, সকল পরিস্থিতিতে মাপকগুলির সুনির্দিষ্ট মান পাওয়া যায় না। কয়েকটি বিচ্ছিন্ন মান প্রদত্ত হলে গড় এবং মধ্যমা সুনির্দিষ্টভাবে নির্ণয় করা সম্ভব ; বিচ্ছিন্ন চলের পরিসংখ্যা-বিভাজনে এক-একটি মান এক-একটি শ্রেণী সূচিত করলে আলোচ্য তিনটি মাপকেরই সাধারণতঃ সুনির্দিষ্ট মান পাওয়া যায়। কিন্তু যদি এক-একটি মানের পরিবর্তে এক-একটি মানসীমা নিয়ে পরিসংখ্যা বিভাজনের শ্রেণীগুলি গঠিত হয় তাহলে বিচ্ছিন্ন বা অবিচ্ছিন্ন চলের ক্ষেত্রে তিনটি মাপকের কোনটিরই সঠিক মান পাওয়া যায় না। স্বল্পসংখ্যক অবিন্যস্ত মান প্রদত্ত হলে বা পরিসংখ্যা-বিভাজনের একাধিক শ্রেণী সর্বোচ্চ পরিসংখ্যা-সম্পন্ন হলে ভূয়িষ্ঠক নির্ণয় প্রত্যক্ষভাবে সম্ভব নয়। গোষ্ঠীবদ্ধ রাশিতথ্যে প্রান্তিক শ্রেণী-দুটির যে কোন একটি অনির্দিষ্ট দৈর্ঘ্যসম্পন্ন হলে ( যেমন, শ্রেণীবিন্যাস যদি এইরকম হয় : 100-এর কম, 101 – 199, 200 – 299, ..., 2500 এবং তদূর্ধ্ব ) গড় নির্ণয় অসম্ভব। অবশ্য এক্ষেত্রে মধ্যমা বা ভূয়িষ্ঠক নির্ণয়ণে কোন অসুবিধা হয় না, যদি না সংশ্লিষ্ট শ্রেণীটি মধ্যমাশ্রেণী বা ভূয়িষ্ঠক-শ্রেণী হয়।

অল্পায়াসে নিরূপণযোগ্যতা আদর্শ মধ্যগামিতা-মাপকের আর একটি প্রয়োজনীয় সর্ত। সাধারণভাবে বলা যায়, তিনটি মাপকই এই সর্তের বিচারে প্রায় সমতুল—তবে গড় নির্ণয় হয়ত অপেক্ষাকৃত সামান্য বেশী শ্রম এবং সময় সাপেক্ষ।

একটি আদর্শ মধ্যগামিতা-মাপকের পক্ষে প্রদত্ত প্রতিটি মানের ওপর প্রত্যক্ষ-ভাবে নির্ভরশীল হওয়া উচিত। একমাত্র গড়নির্ণয়ের ক্ষেত্রেই প্রদত্ত প্রতিটি মান প্রত্যক্ষভাবে গ্রহণ করা হয়ে থাকে, যদিও অন্য দুটি মাপকের মান নির্ধারণে সবকটি মান পরোক্ষভাবে বিবেচনা করা হয়। প্রদত্ত এক বা একাধিক মান পরিবর্তন ক'রেও মধ্যমা বা ভূয়িষ্ঠকের মান অপরিবর্তিত রাখা চলে, কিন্তু গড়ের ক্ষেত্রে সাধারণতঃ তা সম্ভব হয় না।

আদর্শ মধ্যগামিতা-মাপক প্রদত্ত মানগুলির প্রতিনিধি-স্থানীয় হবে এটাই বাঞ্ছনীয়। ভূয়িষ্ঠক এই সর্তের বিচারে সর্বোত্তম, কারণ ভূয়িষ্ঠক সূচিত করে সর্বোচ্চ পরিসংখ্যা-সম্পন্ন মানটি। সাধারণভাবে গড় ও মধ্যমাও সর্তটি পূরণ করে—তবে

প্রদত্ত মানগুলির মধ্যে একটি বা দুটি দলছুট (outlier) মান থাকলে গড় অপেক্ষা মধ্যমার ব্যবহার, এবং প্রদত্ত মানগুলি সম্পূর্ণ ভিন্ন দুটি গোষ্ঠীতে ভাগ হয়ে গেলে মধ্যমা অপেক্ষা গড়ের ব্যবহার, যুক্তিযুক্ত। ধরা যাক, 7 জন পরীক্ষার্থীর প্রাপ্ত নম্বর 80, 72, 68, 5, 69, 76 এবং 92—এক্ষেত্রে স্পষ্টতঃই মধ্যমা (72) গড়মান (66) অপেক্ষা বেশী প্রতিনিধিস্থানীয়, কেননা প্রদত্ত 7টি মানের মধ্যে 6টিই গড়মান অপেক্ষা বৃহত্তর। আবার অনুরূপ উদাহরণে 7টি নম্বর যদি হয় 13, 5, 18, 21, 84, 76, 98—সেক্ষেত্রে অবশ্যই গড়মানটি (45) মধ্যমা (21) অপেক্ষা অধিকতর প্রতিনিধি-স্থানীয়। অবশ্য শেষোক্ত ক্ষেত্রে প্রকৃতপক্ষে মধ্যগামিতার অস্তিত্ব নাই।

আদর্শ মধ্যগামিতা-মাপক **নমুনাজ চাঞ্চল্যের** (sampling fluctuation) যথাসম্ভব কম প্রভাবাধীন হবে, এটাই বাঞ্ছনীয়। নমুনাজ চাঞ্চল্য কথাটি পরবর্তী একটি অধ্যায়ে বিস্তারিতভাবে আলোচিত হবে। তবে একটি ছোট উদাহরণ নিলে এ সম্বন্ধে কিছু ধারণা হতে পারে। মনে কর, একটি শ্রেণীতে ছাত্রসংখ্যা 5 জন—কোন একটি পরীক্ষায় এদের প্রাপ্ত নম্বর যথাক্রমে 51, 75, 57, 72 ও 48. শ্রেণীটি থেকে 3 জন ছাত্রের এক-একটি নমুনা সংগ্রহ ক'রে সম্ভাব্য সবকটি নমুনায় আগত ছাত্রদের প্রাপ্ত নম্বরের গড় ও মধ্যমা নির্ণয় করা হ'ল :

**সারণী 4.5**

**বিভিন্ন নমুনার জন্য নম্বরের গড় ও মধ্যমা নির্ণয়**

| নমুনা | গড় | মধ্যমা |
|---|---|---|
| 51, 75, 57 | 61 | 57 |
| 51, 75, 72 | 66 | 72 |
| 51, 75, 48 | 58 | 51 |
| 51, 57, 72 | 60 | 57 |
| 51, 72, 48 | 57 | 51 |
| 75, 57, 72 | 68 | 72 |
| 75, 57, 48 | 60 | 57 |
| 75, 72, 48 | 65 | 72 |
| 57, 72, 48 | 59 | 57 |
| 51, 57, 48 | 52 | 51 |

কোন একটি মাপকের নমুনাজ চাঞ্চল্য কথাটি সাধারণভাবে নির্দেশ করে ঐ মাপকটির সম্ভাব্য সবকটি নমুনা থেকে পাওয়া মানগুলির পরস্পরের মধ্যে পার্থক্যের গড় পরিমাণ। লক্ষ কর, এখানে গড় অপেক্ষা মধ্যমার নমুনাজ চাঞ্চল্য বেশী। সাধারণভাবে দেখা গেছে, প্রচলিত সবকটি মধ্যগামিতা-মাপকের মধ্যে গড়ের নমুনাজ চাঞ্চল্যই সর্বাপেক্ষা কম। তবে আগের অনুচ্ছেদে যা বলা হয়েছে, সারিতে এক বা একাধিক উল্লেখযোগ্য দলছুট মান থাকলে গড় অপেক্ষা মধ্যমার নমুনাজ চাঞ্চল্য কম হতে পারে।

সুবিধা-অসুবিধাগুলির আপেক্ষিক গুরুত্বের বিচারে সাধারণভাবে গাণিতিক গড় আলোচ্য তিনটি মধ্যগামিতা-মাপকের মধ্যে সর্বশ্রেষ্ঠ বলা চলে। গড়ের আর একটি বড় সুবিধা, এর বিভিন্ন গাণিতিক ধর্ম—যেগুলি পরবর্তী পর্যায়ে বিভিন্ন গাণিতিক পদ্ধতি প্রয়োগে খুবই সহায়ক। অবশ্য 4.4.4 অনুচ্ছেদে আলোচিত মধ্যমার ধর্মটির জন্য যেসব ক্ষেত্রে 'ক্রমের' প্রশ্নটি গুরুত্বপূর্ণ, সেসব ক্ষেত্রে গড় অপেক্ষা মধ্যমার ব্যবহার প্রশস্ত, কারণ গাণিতিক গড়ের এই ধর্মটি নেই। তা ছাড়া, অসমদৈর্ঘ্য শ্রেণীবিন্যাসের ক্ষেত্রে শ্রেণী-অন্তরগুলি সমান না হওয়ার দরুণ গাণিতিক গড় নির্ণয়ে যে অসুবিধা হয়, মধ্যমা বা ভূয়িষ্ঠক নির্ণয়ে সেটি থাকে না।

## 4.8 অন্যান্য মধ্যগামিতা-মাপক :

**4.8.1 গুণোত্তর গড়** (geometric mean) : কোন চলের $x_1, x_2, \ldots, x_n$—এই $n$-টি মান প্রদত্ত হলে চলটির গুণোত্তর গড় $\bar{x}_g$-এর সংজ্ঞা হল

$$\bar{x}_g = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}. \qquad \cdots \quad (4.14a)$$

গোষ্ঠীবদ্ধ রাশিতথ্যের ক্ষেত্রে

$$\bar{x}_g = \left(\prod_{i=1}^{k} x_i^{f_i}\right)^{\frac{1}{n}}, \qquad \cdots \quad (4.14b)$$

যেখানে,
$$n = \sum_{i=1}^{k} f_i.$$

[ এখানে $\prod_{i=1} x_i$ সংকেতচিহ্নটি $x_1 . x_2 \ldots x_k$—এই ধারাবাহিক গুণনের সংক্ষিপ্ত রূপ। ]

লক্ষ্য কর, $\log \bar{x}_g = \begin{cases} \dfrac{1}{n}\sum\limits_{i=1}^{n} \log x_i, & \text{অবিন্যস্ত রাশিতথ্যের ক্ষেত্রে} \quad (4.15a) \\ \dfrac{1}{n}\sum\limits_{i=1}^{k} f_i \log x_i, & \text{গোষ্ঠীবদ্ধ রাশিতথ্যের ক্ষেত্রে} \quad (4.15b) \end{cases}$

অর্থাৎ, প্রদত্ত মানগুলির গুণোত্তর গড়ের লগ, এদের লগারিদমের গাণিতিক গড়ের সমান। স্পষ্টতঃই লগ ব্যবহার ক'রে গুণোত্তর গড় নির্ণয়ের শ্রমসঙ্কোচ করা যায়।

দুটি চলের একাধিক অনুপাতের গড় নির্ণয়ে গুণোত্তর গড়ের ব্যবহার প্রশস্ত। কারণ,

$$\left(\prod_{i=1}^{n} \frac{x_i}{y_i}\right)^{1/n} = \frac{\left(\prod\limits_{i=1}^{n} x_i\right)^{1/n}}{\left(\prod\limits_{i=1}^{n} y_i\right)^{1/n}} \quad \cdots \quad (4.16)$$

অর্থাৎ, দুটি চলের অনুপাতের গুণোত্তর গড় এদের গুণোত্তর গড়ের অনুপাতের সমান। গুণোত্তর গড়ের এই ধর্মটির জন্য **দরের সূচক-সংখ্যা** (index number of prices) নির্ণয়ে বিভিন্ন দ্রব্যের দরের আপেক্ষিকগুলির (price-relative) —অর্থাৎ যে বৎসরে সূচক-সংখ্যা নির্ণয় করা হচ্ছে দ্রব্যটির সেই বৎসরের দর ÷ দ্রব্যটির ভিত্তি বৎসরের দর, এই ভাগফলগুলির, গুণোত্তর গড় নেওয়া হয়ে থাকে। একটা উদাহরণ নেওয়া যেতে পারে। সমান প্রয়োজনীয় দুটি দ্রব্যের একটির দর মনে কর, ভিত্তি বৎসরের তুলনায় বর্তমানে দ্বিগুণ ও অন্যটির অর্ধেক দাঁড়াল। এক্ষেত্রে গড়ে মূল্যমান অপরিবর্তিত থাকার কথা। একমাত্র গুণোত্তর গড় ব্যবহার ক'রেই এই সঠিক চিত্রটি পাওয়া সম্ভব। চক্রবৃদ্ধি সুদের গড় হার, যন্ত্রের অবমূল্যায়নের (depreciation) গড় হার, ইত্যাদি নির্ণয়েও গুণোত্তর গড় ব্যবহার করা হয়ে থাকে।

যথাক্রমে $n_1, n_2, \ldots, n_k$ সংখ্যক মানসম্পন্ন $k$টি গোষ্ঠীর গুণোত্তর গড় $G_1, G_2, \ldots, G_k$ হলে এই $n_1 + \cdots + n_k = n$টি মানের সার্বিক গুণোত্তর গড় $G$-এর

মান হবে $$G = \left(\prod_{i=1}^{k} G_i^{\,n_i}\right)^{1/n}. \quad \cdots \quad (4.17)$$

খুব সহজেই সূত্রটি প্রমাণ করা যায়।

গুণোত্তর গড় ব্যবহারের বিপক্ষে সবথেকে বড় যুক্তি, এটির নির্ণয়ণ প্রচুর শ্রমসাপেক্ষ। তাছাড়া প্রদত্ত মানগুলির যে কোন একটি শূন্য হলেই অন্যান্যগুলি যাই হোক না কেন, গুণোত্তর গড়ের মান শূন্য হবে। আবার এক বা একাধিক মান ঋণাত্মক হলে গুণোত্তর গড়ের মান একটি অবাস্তব সংখ্যা (imaginary number) হতে পারে।

**উদা. 4.11** একব্যক্তি 1,000 টাকা এই সর্তে ধার নিল যে তাকে প্রথম, দ্বিতীয় এবং তৃতীয় বছরে যথাক্রমে শতকরা 4 টাকা, 5 টাকা ও 6 টাকা হারে চক্রবৃদ্ধি সুদ দিতে হবে। দেখা যাক, এই তিনবছরে তার গড় সুদের হার কী দাঁড়ায়।

গড় সুদের হার যদি $r$% হয়, তাহলে স্পষ্টতঃই,

$$1{,}000\left(1+\frac{r}{100}\right)^3 = 1{,}000\left(1+\frac{4}{100}\right)\left(1+\frac{5}{100}\right)\left(1+\frac{6}{100}\right),$$

অর্থাৎ $1+\frac{r}{100}$ হচ্ছে 1·04, 1·05 ও 1·06 এর গুণোত্তর গড়। সুতরাং,

$$\begin{aligned}\log\left(1+\frac{r}{100}\right) &= \tfrac{1}{3}(\log 1{\cdot}04 + \log 1{\cdot}05 + \log 1{\cdot}06)\\ &= \tfrac{1}{3}({\cdot}0170333 + {\cdot}0211893 + {\cdot}0253059)\\ &= {\cdot}0211762 = \log 1{\cdot}04997\end{aligned}$$

অর্থাৎ, $r = 4{\cdot}997$.

### 4.8.2 প্রতিগাণিতিক গড় (harmonic mean) :

$X$ চলের $x_1, x_2, \ldots, x_n$ এই $n$টি মান প্রদত্ত হলে প্রতিগাণিতিক গড় $\bar{x}_h$-এর প্রকাশনটি হবে

$$\bar{x}_h = \frac{n}{\sum_{i=1}^{n} 1/x_i}, \qquad \cdots \quad (4.18a)$$

এবং গোষ্ঠীবদ্ধ রাশিতথ্যের ক্ষেত্রে $\bar{x}_h = \dfrac{n}{\sum_{i=1}^{k} f_i/x_i}$, $\qquad \cdots \quad (4.18b)$

যেখানে $n = \sum_{i=1}^{n} f_i$.

লক্ষ্য কর,
$$\frac{1}{x_h}=\begin{cases}\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\frac{1}{x_i}, & \text{অবিন্যস্ত রাশিতথ্যের ক্ষেত্রে}\\ \dfrac{1}{n}\displaystyle\sum_{i=1}^{k}\frac{f_i}{x_i}, & \text{গোষ্ঠীবদ্ধ রাশিতথ্যের ক্ষেত্রে}\end{cases}$$

অর্থাৎ, প্রদত্ত মানগুলির প্রতিগাণিতিক গড়ের অন্যোন্যক (reciprocal) এদের অন্যোন্যকগুলির গাণিতিক গড়ের সমান। কোন কিছুর 'হার' (rate), যেমন প্রতি ঘণ্টায় গতিবেগ, ইত্যাদির গড় পেতে হলে প্রতিগাণিতিক গড়ই উপযুক্ত মাপক। নীচের উদাহরণটি লক্ষ্য কর।

**উদা. 4.12** একব্যক্তি ঘণ্টায় 60 মাইল বেগে মোটরে কলকাতা থেকে বর্ধমানে পৌঁছে পুনরায় ঘণ্টায় 50 মাইল বেগে কলকাতায় প্রত্যাবর্তন করল। তার গড় গতিবেগ কত ?

মনে করা যাক, কলকাতা থেকে বর্ধমানের দূরত্ব $x$ মাইল। সুতরাং ব্যক্তিটির বর্ধমানে পৌঁছোতে এবং ওখান থেকে কলকাতা ফিরে আসতে সময় লেগেছে যথাক্রমে $x/60$ ঘণ্টা এবং $x/50$ ঘণ্টা। অর্থাৎ, মোট $(x/60+x/50)$ ঘণ্টায় সে $2x$ মাইল অতিক্রম করেছে। সুতরাং তার ঘণ্টায় গড় গতিবেগ

$$2x\Big/\left(\frac{x}{60}+\frac{x}{50}\right)=\ \Big/\left(\frac{1}{60}+\frac{1}{50}\right)=2/(\cdot0167+\cdot0200)=54\cdot5 \text{ মাইল},$$

অর্থাৎ 60 ও 50-এর প্রতিগাণিতিক গড়।

এক্ষেত্রে অতিক্রান্ত মোট দূরত্ব উভয় ক্ষেত্রে সমান। কিন্তু ঘণ্টায় 60 মাইল বেগে 5 ঘণ্টা এবং ঘণ্টায় 50 মাইল বেগে 4 ঘণ্টা—এই 9 ঘণ্টার গড় গতিবেগ গাণিতিক গড়ের সাহায্যেই পাওয়া যাবে।

$x_1, x_2, \ldots, x_n$ যদি কোন চলের $n$-সংখ্যক শূন্যেতর ধনাত্মক মান হয়, তাহলে এগুলির গাণিতিক গড় $(A)$, গুণোত্তর গড় $(G)$ এবং প্রতিগাণিতিক গড় $(H)$-এর মধ্যে নিম্নলিখিত অসমতা-সম্পর্কটি সত্য :

$$A \geqslant G \geqslant H. \qquad \cdots \quad (4.19)$$

**প্রমাণ।** মনে কর, $n=2$.

এখন, $(\sqrt{x_1}-\sqrt{x_2})^2 \geqslant 0$ অর্থাৎ, $x_1+x_2 \geqslant 2\sqrt{x_1x_2}$

অর্থাৎ, $\dfrac{x_1+x_2}{2} \geqslant (x_1x_2)^{\frac{1}{2}}$ অর্থাৎ, $A \geqslant G.$ (4.19a)

আবার $n=4=2^2$ হলে, (4.19a) ফলটি থেকে পাওয়া যায়

$$\frac{1}{2}\left(\frac{x_1+x_2}{2}+\frac{x_3+x_4}{2}\right) > \left\{\frac{x_1+x_2}{2}\cdot\frac{x_3+x_4}{2}\right\}^{\frac{1}{2}}$$

$$> \{\sqrt{x_1x_2}\cdot\sqrt{x_3x_4}\}^{\frac{1}{2}}$$

অর্থাৎ, $\frac{x_1+x_2+x_3+x_4}{4} > (x_1x_2x_3x_4)^{\frac{1}{4}}$.

অর্থাৎ, এক্ষেত্রেও $A > G$. অনুরূপভাবে দেখানো যেতে পারে যে $n=2^k$ ( $k$ যে কোন অখণ্ড ধনসংখ্যা ) হলে $A > G$ সম্পর্কটি সত্য। কিন্তু সাধারণভাবে $n$ এর মান 2 এর কোন গুণিতকের সমান নাও হতে পারে। মনে কর,

$n=2^k-m$ ($m$ ও $k$ অখণ্ড ধনসংখ্যা )।

এখন প্রদত্ত $n$টি মানের সঙ্গে আরও $m$টি নতুন মান নেওয়া যাক। ধর, এই নতুন মানগুলির প্রত্যেকটি $A=\frac{1}{n}\sum_{i=1}^{n} x_i$-এর সমান। সুতরাং এখন মানগুলি দাঁড়াচ্ছে, $x_1, x_2,\ldots, x_n, A, A,\ldots,A$.

যেহেতু এখানে মানগুলির সংখ্যা $n+m=2^k$, সুতরাং

$$\frac{x_1+x_2+\cdots+x_n+mA}{2^k} > \left[x_1x_2\ldots x_n.A^m\right]^{\frac{1}{2^k}}$$

বা, $\frac{nA+mA}{n+m} > \left[G^nA^m\right]^{\frac{1}{2^k}}$

বা, $A^{2^k} > G^nA^m$

বা, $A^{2^k-m} > G^n$, বা, $A^n > G^n$, অর্থাৎ $A > G$. সুতরাং $n$-এর যে কোন মানের জন্য $A > G$ সম্পর্কটি সত্য।

এখন $x_1, x_2, \ldots, x_n$-এদের প্রত্যেকটি যেহেতু ধনাত্মক এবং শূন্যেতর $\frac{1}{x_1}, \frac{1}{x_2}, \ldots, \frac{1}{x_n}$-এদের প্রতিটিও তাই।

সুতরাং আগের ফল অনুসারে $\frac{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}}{n} > \left(\frac{1}{x_1}.\frac{1}{x_2}\ldots\frac{1}{x_n}\right)$

বা, $\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_n}} < (x_1 . x_2\ldots x_n)^{\frac{1}{n}}$

অর্থাৎ, $H < G$.

সুতরাং $A > G > H$ (প্রমাণিত)।

লক্ষ্য কর, (4.19*a*) অসমতা-সম্পর্কটি সমতায় রূপান্তরিত হবে যদি $x_1 = x_2$ হয়। সাধারণভাবে (4.19) অসমতা-সম্পর্কটি সমতায় পর্যবসিত হবে যদি প্রদত্ত প্রতিটি মানই সমান হয়।

### 4.8.3 মধ্যপ্রসার (mid-range) :

চলের ক্ষুদ্রতম এবং বৃহত্তম মানের গাণিতিক গড়কে বলা হয় চলের মধ্য-প্রসার। মধ্যপ্রসারকে অনেক সময় মধ্যগামিতা-মাপক হিসাবে ব্যবহার করা হয়, বিশেষ ক'রে যেসব পরিস্থিতিতে কম সময়ের মধ্যে মধ্যগামিতা মাপনার প্রয়োজন থাকে। **রাশিবিজ্ঞান-সম্মত গুণনিয়ন্ত্রণের** (statistical quality control) ক্ষেত্রে এই মাপকটির ব্যবহার রয়েছে।

সারণী 3.6 থেকে দৈনন্দিন সর্বোচ্চ তাপমাত্রার মধ্যপ্রসারের মান পাওয়া যায় $\frac{1}{2}$ (107·6 + 83·2) ডিঃ ফাঃ = 95·4 ডিঃ ফাঃ।

## 4.9 ভারযুক্ত গড় (weighted average) :

অনেক সময় দেখা যায়, যে রাশিগুলির গড় নির্ণয় করা প্রয়োজন সেগুলি সমান গুরুত্বপূর্ণ নয়। সেক্ষেত্রে রাশিগুলির গড় মানে প্রতিটি রাশির আপেক্ষিক গুরুত্ব বা ভারকে যথাযথভাবে প্রতিফলিত করার উদ্দেশ্যে ভারযুক্ত গড়ের কথা বলা হয়েছে। $x_1, x_2, \ldots, x_n$—প্রদত্ত এই $n$টি রাশি যথাক্রমে $w_1, w_2, \ldots, w_n$ পরিমাণ গুরুত্বসম্পন্ন বা ভারসম্পন্ন হলে রাশিগুলির ভারযুক্ত গাণিতিক গড় $A_w$, গুণোত্তর গড় $G_w$ এবং প্রতিগাণিতিক গড় $H_w$-এর প্রকাশন হবে যথাক্রমে

$$A_w = \sum_{i=1}^{n} w_i x_i \Big/ \sum_{i=1}^{n} w_i, \tag{4.20}$$

$$G_w = \left(\prod_{i=1}^{n} x_i^{w}\right)^{1 \Big/ \sum_{i=1}^{n} w_i}, \tag{4.21}$$

এবং $$H_w = \frac{\sum_{i=1}^{n} w_i}{\sum_{i=1}^{n} w_i/x_i}. \tag{4.22}$$

লক্ষ্য কর, গোষ্ঠীবদ্ধ রাশিতথ্যের ক্ষেত্রে $A$, $G$ এবং $H$ [ যথাক্রমে (4.2), (4.14*b*) এবং (4.18*b*) সূত্রে ] এক ধরনের ভারযুক্ত গড়। এক্ষেত্রে বিভিন্ন পরিসংখ্যাই সংশ্লিষ্ট শ্রেণী-মধ্যকের ভার, এবং $n = \sum_i f_i$ = মোট ভার।

অসম গুরুত্বসম্পন্ন বিভিন্ন গোষ্ঠীর সূচক-সংখ্যাগুলি একত্রিত ক'রে সার্বিক সূচক-সংখ্যা ভারযুক্ত গড়ের সাহায্যে নিরূপণ করা হয়। নীচের উদাহরণটি লক্ষ্য কর।

**উদা. 4.13** নীচের সারণীতে পশ্চিমবঙ্গে 1966-67 সালে বিভিন্ন ধরনের কৃষিজ পণ্যের উৎপাদনের সূচক-সংখ্যা এবং আপেক্ষিক গুরুত্ব দেওয়া হয়েছে [ ভিত্তিকাল = শস্য-বৎসর 1949-50 ]:

**সারণী 4.6**

| কৃষিপণ্যের ধরন | আপেক্ষিক গুরুত্ব $w_i$ | সূচক $I_i$ |
|---|---|---|
| খাদ্য-শস্য | 73·61 | 119·97 |
| তৈলবীজ | 1·27 | 116·52 |
| পাট | 9·09 | 183·77 |
| চা | 9·22 | 120·93 |
| বিবিধ | 6·81 | 150·31 |
| মোট | 100·00 | – |

এখানে ভারযুক্ত গাণিতিক গড়ের সাহায্যে সার্বিক সূচক-সংখ্যা নির্ণয় করা যাক। সার্বিক সূচক

$$I = \sum_{i=1}^{5} I_i w_i \Big/ \sum_{i=1}^{5} w_i$$

$$= 127 \cdot 44.$$

## 4.10 অনুশীলনী

4.1 পরিসংখ্যা-বিভাজনের মধ্যগামিতা কথাটির অর্থ কী? কয়েকটি বহুল ব্যবহৃত মধ্যগামিতা-মাপকের সংজ্ঞা দাও।

4.2 আদর্শ মধ্যগামিতা-মাপক হিসাবে গাণিতিক গড়, মধ্যমা ও ভূয়িষ্ঠকের তুলনামূলক আলোচনা কর।

4.3 নিম্নলিখিত চলগুলির ক্ষেত্রে তুমি মধ্যগামিতা মাপনার জন্য কোন্ কোন্ মাপক ব্যবহারের পক্ষপাতী যুক্তি-সহকারে নির্দেশ কর :

(i) মাথার খুলির মাপ; (ii) ব্যক্তিগত মালিকানায় জমির পরিমাণ; (iii) পরীক্ষায় প্রাপ্ত নম্বর; (iv) ব্যক্তিগত আয় ( পরিসংখ্যা-বিভাজনে প্রদত্ত ); (v) বিভিন্ন লটে ক্রীত দ্রব্যের দর; (vi) দরের বৃদ্ধিহার; (vii) গতিবেগ; (viii) যন্ত্রের অবমূল্যায়নের হার।

4.4 গাণিতিক গড়, গুণোত্তর গড় এবং প্রতিগাণিতিক গড়ের সংজ্ঞা দাও। শেষোক্ত গড় দুটি কোন্ কোন্ পরিস্থিতিতে ব্যবহার করা প্রশস্ত? গড় তিনটির মধ্যে অসমতা-সম্পর্কটি প্রমাণ কর। সম্পর্কটি কখন সমতায় পর্যবসিত হবে?

4.5 (*a*) ক ও খ যথাক্রমে কলকাতা ও ধানবাদ থেকে মোটরসাইকেলে পরস্পরের অভিমুখে যাত্রা শুরু করল। সমস্ত পথের প্রথম, দ্বিতীয় ও তৃতীয় চতুর্থাংশ ক অতিক্রম করল ঘণ্টায় যথাক্রমে 30, 36 ও 40 কি.মি. বেগে। এদিকে খ-এর পথে অতিক্রান্ত মোট সময়ের প্রথম, দ্বিতীয় ও তৃতীয় চতুর্থাংশে গতিবেগ দাঁড়াল ঘণ্টায় যথাক্রমে 32, 34 ও 42 কি.মি.। উভয়েই গড়ে ঘণ্টায় 35 কি.মি. বেগে সমস্ত পথ অতিক্রম করলে পথের শেষ চতুর্থাংশে ক-এর, এবং মোট ভ্রমণ সময়ের শেষ চতুর্থাংশে খ-এর গতিবেগ নির্ণয় কর।

(*b*) নিম্নলিখিত রাশিগুলির গাণিতিক গড়, গুণোত্তর গড়, ও প্রতিগাণিতিক গড় নির্ণয় কর : 1296, 215·1, 17·24, 2·912, ·9235 এবং ·01642.

4.6 (*a*) মনে কর, একটি চলের $x_1, x_2, \ldots, x_n$—এই $n$টি মান দেওয়া আছে। মানগুলির প্রত্যেকটি একই 'পরিমাণে' বাড়ানো হলে (i) গাণিতিক গড়, (ii) মধ্যমা, (iii) ভূয়িষ্ঠক এবং (iv) গুণোত্তর গড় কি-ভাবে পরিবর্তিত হবে? প্রত্যেকটি একই 'অনুপাতে' বাড়ানো হলেই বা এইসব মাপক কি-ভাবে পরিবর্তিত হবে?

(*b*) ঋজুরৈখিক রূপান্তরের (linear transformation) ক্ষেত্রে মূল চলের এবং রূপান্তরিত চলের (i) গাণিতিক গড়, (ii) মধ্যমা এবং (iii) ভূয়িষ্ঠকের মধ্যে কী ধরনের সম্পর্ক থাকবে?

(*c*) 100টি বৃত্তাকার ধাতব পাতের ব্যাসার্ধগুলির গড় এবং মধ্যমা যথাক্রমে 0·25 মিটার ও 0·23 মিটার জানা থাকলে ধাতব পাতগুলির আয়তনের গড় ও মধ্যমা সম্বন্ধে কতখানি বলা যায় ?

(*d*) একটি ঘড়ি প্রতিদিন সকাল 6 টায় ( সঠিক সময় ) মেলানোর সময় লক্ষ্য করা হয় ঘড়িটির প্রদর্শিত সময় 5টা 55 থেকে 6 টা 5-এর মধ্যে থাকে। গত একমাসে ঘড়িটির প্রদর্শিত সময়ের গাণিতিক গড় ও মধ্যমা যথাক্রমে 6টা 1 মিঃ ও 6 টা 0 মিঃ 30 সেঃ হলে এই একমাসে ঘড়িটির ভ্রান্তির পরিমাণের ( মন্দা এবং দ্রুতি যে কোন দিকে ) গড় ও মধ্যমা সম্বন্ধে কতখানি বলা যায় ?

4.7 যদি $n_1$টি মানের গাণিতিক গড় ও মধ্যমা যথাক্রমে $\bar{x}_1$ ও $M_1$ হয় এবং অন্য $n_2$টি মানের গাণিতিক গড় ও মধ্যমা হয় যথাক্রমে $\bar{x}_2$ ও $M_2$, তাহলে দেখাও যে একত্রে এই $n_1+n_2$টি মানের গড় $\bar{x}$ ও মধ্যমা $M$ যথাক্রমে $\bar{x}_1$ এবং $\bar{x}_2$ ও $M_1$ এবং $M_2$-এর মধ্যে অবস্থান করবে।

4.8 একটি পরিসংখ্যা-বিভাজনে শ্রেণীগুলির ঊর্ধ্বসীমান্ত এবং অধঃসীমান্তের অনুপাত একটি ধ্রুবক, ধরা যাক $e^c$. বিভাজনটির গুণোত্তর গড় $=G$, $i$-তম শ্রেণীর পরিসংখ্যা $=f_i$ এবং $i$-তম শ্রেণী-মধ্যক $=e^{x_i}$ হলে প্রমাণ কর,

$$\log_e G = x_1 + \frac{c}{n} \sum_{i=1}^{k} f_i(i-1), \left[n = \sum_{i=1}^{k} f_i\right].$$

4.9 কোনও চল $a, ar, \ldots, ar^{n-1}$ এই $n$টি মান পরিগ্রহণ করে সমান সংখ্যায়। চলটির গাণিতিক গড়, গুণোত্তর গড় ও প্রতিগাণিতিক গড় নির্ণয় কর। দেখাও যে এক্ষেত্রে $AH=G^2$.

4.10 নীচে 1972 সালের কলকাতা প্রথম ডিভিশন ফুটবল লীগের সর্বশেষ অবস্থা দেওয়া হ'ল :

| দল | খেলা | জয় | ড্র | পরাজয় | গোলসংখ্যা | | পয়েন্ট |
|---|---|---|---|---|---|---|---|
| | | | | | স্বপক্ষে | বিপক্ষে | |
| ইস্টবেঙ্গল | 19 | 18 | 1 | 0 | 44 | 0 | 37 |
| মোহনবাগান | 19 | 16 | 2 | 1 | 50 | 8 | 34 |
| মহঃ স্পোর্টিং | 19 | 15 | 2 | 2 | 36 | 8 | 32 |
| এরিয়ান্স | 19 | 8 | 5 | 6 | 26 | 14 | 21 |
| বি.এন.আর | 19 | 7 | 7 | 5 | 23 | 19 | 21 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| উয়াড়ী | 19 | 6 | 8 | 5 | 19 | 18 | 20 |
| রাজস্থান | 19 | 8 | 4 | 6 | 15 | 14 | 20 |
| ভ্রাতৃসঙ্ঘ | 19 | 6 | 7 | 6 | 15 | 12 | 19 |
| খিদিরপুর | 19 | 6 | 7 | 6 | 12 | 11 | 19 |
| বাটা | 19 | 6 | 5 | 8 | 16 | 19 | 17 |
| পোর্ট কমিঃ | 19 | 7 | 2 | 10 | 20 | 26 | 16 |
| জর্জ টেলিঃ | 19 | 5 | 5 | 9 | 16 | 25 | 15 |
| ক্যাল্ জিম্খানা | 19 | 5 | 5 | 9 | 15 | 26 | 15 |
| হাওড়া ইউঃ | 19 | 4 | 7 | 8 | 13 | 17 | 15 |
| কালীঘাট | 19 | 5 | 4 | 10 | 14 | 26 | 14 |
| টালি অগ্রঃ | 19 | 4 | 6 | 9 | 13 | 20 | 14 |
| বালি প্রতিভা | 19 | 4 | 6 | 9 | 9 | 31 | 14 |
| কুমারটুলি | 19 | 3 | 8 | 8 | 7 | 25 | 14 |
| ইঃ রেল | 19 | 3 | 6 | 10 | 5 | 26 | 12 |
| স্পোর্টিং ইউঃ | 19 | 4 | 3 | 12 | 4 | 26 | 11 |

[ উৎস : আনন্দবাজার ৯।৫।৭৩ ]

(i) দলগুলির জয়ের সংখ্যা ($x_1$), ড্র-এর সংখ্যা ($x_2$), পরাজয়ের সংখ্যা ($x_3$), স্বপক্ষে গোলের সংখ্যা ($x_4$) ও বিপক্ষে গোলের সংখ্যা ($x_5$) এবং পয়েণ্ট সংখ্যার ($y$) গাণিতিক গড় ও মধ্যমা বের কর।

(ii) প্রত্যক্ষ কর, $\bar{x}_1 = \bar{x}_3$ এবং $\bar{x}_4 = \bar{x}_5$. এরূপ হওয়ার কারণ কী ? মধ্যমার ক্ষেত্রেও কি অনুরূপ সম্পর্ক সত্য ?

(iii) $\bar{x}_1$-এর মান জানা থাকলে $\bar{x}_2$-এর মান বের করা যায় কি ? কি-ভাবে ? অনুরূপভাবে $\tilde{x}_1$-এর মান থেকে $\tilde{x}_2$-এর মান পাওয়া কি সম্ভব ? কেন ?

(iv) প্রতিটি জয়ের জন্য 2 পয়েণ্ট এবং প্রতিটি ড্র-এর জন্য 1 পয়েণ্ট হিসেবে বিভিন্ন দলের পয়েণ্ট নির্ধারণ করা হয়। সেক্ষেত্রে $\bar{x}_1$-এর মান থেকে $\bar{y}$-এর মান পাওয়া সম্ভব কি ? অনুরূপভাবে কি $\tilde{y}$-এর মান নির্ণয় করা যায় ?

(v) (i)-(iv) প্রশ্নাংশে 19টি খেলায় জয়, ড্র, ···ইত্যাদি সংখ্যার দলপ্রতি গড় ও মধ্যমা চাওয়া হয়েছে। এই ফলগুলি ব্যবহার ক'রে দলগুলির খেলাপ্রতি (per team per game) জয়, ড্র, ···ইত্যাদি সংখ্যার গড় ও মধ্যমা নির্ণয় কর।

4.11 3.7 অনুশীলনীতে প্রদত্ত রাশিতথ্যের পরিসংখ্যা-বিভাজনের জন্য বিভিন্ন মধ্যগামিতা-মাপক, প্রথম ও তৃতীয় চতুর্থক এবং চতুর্থ ও ষষ্ঠ দশমকের মান নির্ণয় কর।

4.12 3.8 অনুশীলনীতে প্রদত্ত রাশিতথ্যের জন্য 4.11 অনুশীলনীতে উল্লেখিত বিবরণাত্মক মাপকগুলির মান নির্ণয় কর।

4.13 নীচের সারণীতে প্রদত্ত পরিসংখ্যা-বিভাজনের গাণিতিক গড় নির্ণয় কর।

**সারণী 4.7**

**1931 সালে জয়পুর শহরে পুরুষ অধিবাসীদের পরিসংখ্যা-বিভাজন**

| বয়স | জনসংখ্যা ( হাজারে ) |
|---|---|
| 0— 5 | 9 |
| 5—10 | 8 |
| 10—15 | 8 |
| 15—20 | 7 |
| 20—30 | 15 |
| 30—40 | 12 |
| 40—50 | 9 |
| 50—60 | 6 |
| 60—70 | 4 |
| মোট | 78 |

[ ইঙ্গিত : এখানে শ্রেণীবিন্যাস সমদৈর্ঘ্য নয় শ্রেণীদৈর্ঘ্যগুলি 5-এর গুণিতক। $y = x - A$ নাও। ]

4.14(*a*) নীচের পরিসংখ্যা-বিভাজন সারণীতে দুটি পরিসংখ্যা দেওয়া নাই। বিভাজনটির গাণিতিক গড় = 49·65 কি.গ্রা. জানা আছে। অপ্রদত্ত পরিসংখ্যা দুটি নির্ণয় কর :

**সারণী 4.8**

**100 জন ছাত্রের ওজনের পরিসংখ্যা-বিভাজন**

| ওজন ( কি.গ্রা. ) | পরিসংখ্যা |
|---|---|
| 35·0—39·9 | 5 |
| 40·0—44·9 | * |
| 45·0—49·9 | 30 |
| 50·0—54·9 | 23 |
| 55·0—59·9 | * |
| 60·0—64·9 | 8 |
| 65·0—69·9 | 1 |
| মোট | 100 |

[ ইঙ্গিত : মনে কর, অপ্রদত্ত পরিসংখ্যা দুটি যথাক্রমে $f_1$ এবং $f_2$. $u=\frac{1}{5}(x-57{\cdot}45)$ লিখে পাওয়া যায়

$$49{\cdot}65 = 57{\cdot}45 + \tfrac{5}{100}\{8\times1+1\times2-23\times1-30\times2-3f_1-4\times5\}$$

এবং $f_1+f_2=100-(5+30+23+8+1)$. ]

(*b*) গাণিতিক গড়ের পরিবর্তে মনে কর মধ্যমার মান জানা আছে 48·95 কি.গ্রা.। সেক্ষেত্রে অপ্রদত্ত পরিসংখ্যা দুটি কি-ভাবে বের করবে ?

4.15 নীচের সারণীতে একটি পরিসংখ্যা-বিভাজনের বিভিন্ন দশমকের মানক্রমগুলি দেওয়া হয়েছে। আয়তলেখটি অঙ্কিত কর।

| দশমক | 0% | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|---|---|---|---|---|---|---|---|---|---|---|---|
| দশমকের মানক্রম | 0 | 4 | 8 | 18 | 49 | 64 | 82 | 91 | 95 | 99 | 100 |

4.16 একটি কারখানার 377 জন শ্রমিকের মজুরি সংক্রান্ত নিম্নলিখিত তথ্য পাওয়া গেছে। কারখানায় প্রতি শতে মজুরির হার নির্ণয় কর।

| শ্রমিক পিছু দৈনিক উৎপাদন | শ্রমিক-সংখ্যা | প্রতি শতে মজুরির হার ( টাকায় ) |
|---|---|---|
| 401—500 | 32 | 0·75 |
| 501—600 | 79 | 1·00 |
| 601—700 | 142 | 1·25 |
| 701—800 | 89 | 1·50 |
| 801—900 | 31 | 1·75 |
| 901—1000 | 4 | 2·00 |
| মোট | 377 | — |

## 4.11 নির্দেশিকা

1. Cook, L.H.L. *Statistical Problems and How to Solve Them.* Barnes & Noble, 1971.

2. Goon, A.M., Gupta M.K. & Dasgupta B. *Fundamentals of Statistics*, Vol. I. World Press, 1975.

3. Kenney, J.F. & Keeping, E.S. *Mathematics of Statistics*, Part I. Van Nostrand, 1954.

4. Mounse, J. *Introduction to Statistical Calculations.* English University Press, 1952.

5. Simpson, G. & Kafka, F. *Basic Statistics.* H. Holt, 1955.

6. Yule, G.U. & Kendall, M.G. *An Introduction to the Theory of Statistics*. Charles Griffin, 1968.

# বিস্তৃতি এবং বিস্তৃতি-মাপক

## ( Dispersion and its Measures ) 5

### 5.1 বিস্তৃতি (dispersion) কী ?

ভালোভাবে লক্ষ্য করলে দেখা যাবে, কোন চল কর্তৃক পরিগৃহীত কিছু মানের মধ্যগামিতাই একমাত্র বৈশিষ্ট্য নয়, চলটির বিভাজনের মধ্যগামিতা সম্পর্কে ধারণা পাওয়ার পরও বিভাজনটির প্রকৃতি সম্বন্ধে আরও কিছু জানার থাকে। যেমন, গ্রীষ্মকালে দামোদরে গড়ে একহাঁটু জল থাকে, সুতরাং ঐ সময় নদীর উৎসমুখ থেকে মোহনা পর্যন্ত হেঁটেই নদীটি পার হওয়া যাবে এইরকম সিদ্ধান্ত যদি কেউ নিয়ে বসেন, তিনি অবশ্যই বিপদে পড়বেন। কারণ স্পষ্টতঃই বিভিন্ন স্থানে নদীটির গভীরতা কতখানি বাড়ে বা কমে তা বিস্তারিতভাবে জানা প্রয়োজন এই প্রসঙ্গে। বস্তুতঃ একটি নির্বাচিত মধ্যগামিতা-মাপকের মানটি প্রদত্ত একপ্রস্থ রাশিতথ্যের প্রতিভূ হিসাবে ব্যবহার করা যেতে পারে, কিন্তু চলের যে প্রধান ধর্ম পরিবর্তনশীলতা তার প্রভাবে চলটির বিভিন্ন মানগুলির একটি অপরটি থেকে স্বভাবতঃই ভিন্ন—ফলে দুপ্রস্থ রাশিতথ্যের গড়মান অভিন্ন হয়েও এদের বিভাজনের মধ্যে প্রকৃতিগত বৈসাদৃশ্য থাকা সম্ভব। একটি উদাহরণ নিলে বক্তব্যটি পরিষ্কার হবে। ধরা যাক, ভারতের জাতীয় দলে প্রতিনিধিত্ব করার দাবীদার দুজন প্রতিযোগী ক্রিকেটারের প্রথম শ্রেণীর ক্রিকেটে পর পর দশটি ইনিংসে সংগৃহীত রানের সংখ্যা এই রকম :

| ইনিংস সংখ্যা | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| ক-এর রান | 62 | 63 | 64 | 55 | 62 | 56 | 58 | 62 | 60 | 58 |
| খ-এর রান | 119 | 23 | 79 | 0 | 68 | 0 | 42 | 85 | 19 | 165 |

গাণিতিক গড়ের বিচারে এই দুপ্রস্থ রানসংখ্যার মধ্যগামিতা অভিন্ন—উভয় ক্ষেত্রেই গড় 60। তবুও এই দুপ্রস্থ রানসংখ্যার প্রকৃতি কিন্তু এক নয়। ক-এর রানসংখ্যা সবগুলি ইনিংসেই মোটামুটি 60-এর কাছাকাছি। কিন্তু খ-এর সংগৃহীত রান একদিকে যেমন 165 তে পৌঁছেছে, তেমনি দু-দুবার তাকে প্যাভেলিয়নে ফিরতে হয়েছে শূন্যহাতে। তাই খ-এর কৃতিত্বে দুটি সেঞ্চুরী

থাকলেও ক-এর ব্যাটিং যেহেতু অনেক বেশী সমঞ্জস (consistent), তাই নির্বাচকমণ্ডলীর ক-কেই বেশী পছন্দ হওয়ার কথা।

সুতরাং দেখা যাচ্ছে, চল কর্তৃক পরিগৃহীত মানগুলির যেমন একটি কেন্দ্রীয় মানের দিকে কেন্দ্রীভূত হওয়ার প্রবণতা রয়েছে, তেমনি এই কেন্দ্রীয় মানের উভয়পার্শ্বে মানগুলি কতদূর বিস্তৃত এবং পরস্পরের মধ্যে কী ধরনের পার্থক্য রয়েছে—অর্থাৎ মানগুলি কি-ভাবে ছড়িয়ে রয়েছে—সে ব্যাপারেও চলটির কিছু নিজস্ব বৈশিষ্ট্য থাকা সম্ভব। চলের এই বৈশিষ্ট্যটিকেই রাশিবিজ্ঞানের পরিভাষায় **বিস্তৃতি** (dispersion) বলে। এক কথায় চলের বিস্তৃতি হ'ল এর মানগুলির বিক্ষেপণের মাত্রা (degree of scatter)। ওপরের উদাহরণ থেকে বোঝা যাচ্ছে, দুটি চলের মধ্যে বা দুগ্রস্থ রাশিতথ্যের মধ্যে সার্থক তুলনা করতে হলে মধ্যগামিতার পাশাপাশি বিস্তৃতির বিচারও করা প্রয়োজন।

তিন ধরনের বিস্তৃতি মাপকের প্রচলন রয়েছে। এক, প্রান্তিক মানগুলির অথবা ভগ্নাংশকের ভিত্তিতে—যেমন, **প্রসার** (range), **চতুর্থক বিচ্যুতি** বা **আন্তঃচতুর্থক অর্ধপ্রসার** (quartile deviation or, semi-interquartile range); দুই, কোন কেন্দ্রীয় মানের সঙ্গে পার্থক্যের ভিত্তিতে—যেমন, **গড় বিচ্যুতি** (mean deviation), **মূল-গড়-বর্গ বিচ্যুতি** (root-mean-square deviation), **প্রমাণ বিচ্যুতি** (standard deviation); তিন, গৃহীত মান-গুলির পারস্পরিক পার্থক্যের ভিত্তিতে—যেমন, **গিনির গড় পার্থক্য** (Gini's mean difference)। তবে প্রসার, গড়বিচ্যুতি এবং প্রমাণবিচ্যুতি—এই তিনটিই হ'ল অপেক্ষাকৃত বহুল ব্যবহৃত বিস্তৃতি-মাপক। পরবর্তী কয়েকটি অনুচ্ছেদে বিভিন্ন বিস্তৃতি-মাপক সম্বন্ধে বিস্তারিত আলোচনা করা হয়েছে।

### 5.2 প্রসার (range) :

চলের সংগৃহীত মানগুলির মধ্যে সর্বোচ্চ ও সর্বনিম্ন মানের পার্থক্যের পরিমাণকে চলটির প্রসার বলা হয়। 5.1 অনুচ্ছেদে প্রদত্ত উদাহরণে ক ও খ-এর রানসংখ্যার প্রসার যথাক্রমে $65-55=10$ এবং $165-0=165$.

যদিও এটি প্রদত্ত সবকটি মানের ওপর প্রত্যক্ষভাবে নির্ভরশীল নয়, তবুও খুব সহজে নির্ণয় করা যায় ব'লে বিস্তৃতি-মাপক হিসাবে প্রসারের যথেষ্ট প্রচলন আছে। লক্ষণীয়, শ্রেণীবিন্যস্ত রাশিতথ্যে—যেখানে এক-একটি শ্রেণী এক-একটি শ্রেণী-অন্তর সূচিত করে—শেষ শ্রেণীর ঊর্ধ্বসীমান্ত থেকে প্রথম শ্রেণীর অধঃসীমান্তের বিয়োগফলই বিভাজনটির প্রসার।

3.3 ও 3.6 সারণীতে পরিবেশিত রাশিতথ্যের প্রসার যথাক্রমে $6-0=6$ এবং $107{\cdot}6-83{\cdot}2=24{\cdot}4$ (উপযুক্ত এককে)। শেষোক্ত ক্ষেত্রে সংশ্লিষ্ট পরিসংখ্যা-বিভাজন থেকে পাওয়া প্রসারের মান $108{\cdot}95-81{\cdot}95=27{\cdot}0$ (উপযুক্ত এককে)। অবিন্যস্ত রাশিতথ্য থেকে পাওয়া প্রসারের মান যে সংশ্লিষ্ট গোষ্ঠীবদ্ধ রাশিতথ্য থেকে পাওয়া প্রসারের মান অপেক্ষা কিছু কম হবে, তা সহজেই বোঝা যায়।

**5.3 চতুর্থক বিচ্যুতি** (quartile deviation) **বা আন্তঃ-চতুর্থক অর্ধপ্রসার** (semi-interquartile range) **:**

চতুর্থ পরিচ্ছেদে পরিসংখ্যা-বিভাজনের ভগ্নাংশক এবং ভগ্নাংশকের বিশেষ বিশেষ রূপ, যথা—মধ্যমা, চতুর্থক, ইত্যাদির সংজ্ঞা আলোচিত হয়েছে। এখন স্পষ্টতঃই কোন চল কর্তৃক গৃহীত মানগুলি যত বেশী বিস্তৃত হবে, এর প্রথম ও তৃতীয় চতুর্থকের পার্থক্যও হবে তত বেশী। তাই

$$Q=\tfrac{1}{2}(Q_3-Q_1) \qquad \cdots \qquad \cdots \qquad (5.1)$$

এই রাশিটিকে একটি বিস্তৃতি-মাপক হিসাবে ব্যবহার করা হয়। এটিই হ'ল চতুর্থক বিচ্যুতি বা আন্তঃচতুর্থক অর্ধপ্রসার।

এক্ষেত্রেও মাপকটি প্রদত্ত সবকটি মানের ওপর প্রত্যক্ষভাবে নির্ভরশীল নয়। তবে গোষ্ঠীবদ্ধ রাশিতথ্যে প্রান্তিক শ্রেণী-দুটির যে-কোনটি অনির্দিষ্ট সীমা-সম্পন্ন হলে, (যেমন 10 অথবা তার কম, 11—20, 21—30, ..., 91 এবং তদূর্ধ্ব—এই ধরনের শ্রেণীবিন্যাসে) বিস্তৃতি পরিমাপনে এই মাপকটি অপরিহার্য। অসমদৈর্ঘ্য শ্রেণীবিন্যাসেও এই মাপকটি খুবই উপযোগী।

**উদা. 5.1** 3.5 এবং 3.9 সারণীতে প্রদত্ত রাশিতথ্যের ক্ষেত্রে চতুর্থক বিচ্যুতি বের করা যাক।

প্রথম উদাহরণে $Q_1=1$, $Q_3=2$. সুতরাং $Q=\tfrac{1}{2}(2-1)={\cdot}5$.

দ্বিতীয় উদাহরণে, $Q_3=99{\cdot}2288$ এবং $Q_1=92{\cdot}0387$.

সুতরাং $Q=\tfrac{1}{2}(99{\cdot}2288-92{\cdot}0387)$ ডিঃ ফাঃ

$=7{\cdot}1981/2$ ডিঃ ফাঃ $=3{\cdot}5991$ ডিঃ ফাঃ।

**5.4 গড় বিচ্যুতি** (mean deviation) **:**

ওপরের দুটি অনুচ্ছেদে আলোচিত বিস্তৃতি-মাপক দুটি প্রান্তিক মান অথবা ভগ্নাংশক-ভিত্তিক। আগেই বলা হয়েছে, মাপক-দুটি সংশ্লিষ্ট প্রতিটি মানের ওপর প্রত্যক্ষভাবে নির্ভরশীল নয়। সংশ্লিষ্ট প্রতিটি মানের ওপর প্রত্যক্ষভাবে

নির্ভরশীল একটি বিস্তৃতি-মাপক হ'ল গড়বিচ্যুতি। একটি চলের $x_1, x_2, \ldots, x_n$ —এই $n$টি বিচ্ছিন্ন মান দেওয়া আছে ধরা যাক। $A$ যদি একটি যথেচ্ছ-গৃহীত মান হয় তাহলে সংজ্ঞানুসারে প্রদত্ত মানগুলির $A$-কেন্দ্রিক গড়বিচ্যুতি হ'ল

$$MD_A = \frac{1}{n}\sum_{i=1}^{n} |x_i - A|. \quad \cdots \quad (5.2)$$

শ্রেণীবিন্যস্ত রাশিতথ্যের ক্ষেত্রে

$$MD_A = \frac{1}{n}\sum_{i=1}^{k} f_i\,|x_i - A|, \quad \cdots \quad (5.3)$$

যেখানে, $x_i = i$-তম শ্রেণীর যথার্থ মান অথবা মধ্যক,

$f_i = i$-তম শ্রেণীর পরিসংখ্যা,

এবং $n = \sum_{i=1}^{k} f_i.$

এখানে লক্ষণীয়, $|x|$ চিহ্নটির অর্থ হ'ল, $x$-এর চিহ্ন-নিরপেক্ষ, অর্থাৎ পরম মান (absolute value)। যেমন $|5| = 5$, $|-5| = 5$.

এখন বিস্তৃতি-মাপকের কাজ সাধারণতঃ সংশ্লিষ্ট মানগুলি একটি কেন্দ্রীয় মানের উভয়পাশে কী পরিমান বিস্তৃত তা পরিমাপ করা—তাই গড়বিচ্যুতি নির্ণয়ের সময় যথেচ্ছ-গৃহীত বিন্দু $A$টি সাধারণতঃ নেওয়া হয় কোনও মধ্যগামিতা-মাপকের মান, যেমন গাণিতিক গড়, মধ্যমা বা ভূয়িষ্ঠক। গড়বিচ্যুতিকেও সেইমত বলা হয় গড়কেন্দ্রিক, মধ্যমাকেন্দ্রিক বা ভূয়িষ্ঠককেন্দ্রিক গড়বিচ্যুতি। কেন্দ্র সম্বন্ধে কোন উল্লেখ না থাকলে সাধারণতঃ গড়কেন্দ্রিক গড়বিচ্যুতিই নেওয়া হয়ে থাকে।

স্পষ্টতঃই $i$ এর বিভিন্ন মানের জন্য $|x_i - A|$ এর পরিমান যত বেশী, চলটির বিস্তৃতিও তত বেশী। এখন প্রশ্ন উঠতে পারে চিহ্নযুক্ত বিচ্যুতি $(x_i - A)$ এর পরিবর্তে বিচ্যুতিগুলির পরমমান $|x_i - A|$ নেওয়া হ'ল কেন। এর উত্তর হ'ল, চিহ্নযুক্ত বিচ্যুতিগুলির যোগফল $\sum_{1}^{n} (x_i - A)$-এ $A$ যদি কোন মধ্যগামিতা-মাপকের মান হয়, তাহলে এতে স্পষ্টতঃই ধনাত্মক বিচ্যুতিগুলির যোগফল এবং ঋণাত্মক বিচ্যুতিগুলির যোগফল প্রায় সমান সমান হবে এবং ফলে, আলাদা আলাদা ভাবে বিচ্যুতিগুলির মান উল্লেখযোগ্য হলেও মোট যোগফলটির মান হবে শূন্যের খুবকাছাকাছি। আর $A$ যদি গাণিতিক গড় হয়, তাহলে যোগফলটি তো গাণিতিক

গড়ের ধর্ম অনুযায়ী যথার্থই শূন্য। সুতরাং $\frac{1}{n}\sum_{i=1}^{n}(x_i - A)$ মাপকটি সংশ্লিষ্ট মানগুলির বিস্তৃতির প্রকৃত চিত্র দিতে সক্ষম হয় না। চিহ্ন-নিরপেক্ষ বিচ্যুতি নেওয়ার ফলে এই অসুবিধাটি দূর হয়েছে।

প্রমাণ করা যেতে পারে, মধ্যমাকেন্দ্রিক গড়বিচ্যুতি অন্য যে কোন বিন্দু-কেন্দ্রিক গড়বিচ্যুতি অপেক্ষা বৃহত্তর হতে পারে না (অনু 5.7)। তাই অনেকে বিস্তৃতি-মাপক হিসাবে মধ্যমাকেন্দ্রিক গড়বিচ্যুতি ব্যবহারের পক্ষপাতী।

**উদা. 5.2** 5.1 অনুচ্ছেদে বর্ণিত উদাহরণে ক এবং খ এর রানসংখ্যার গড়-কেন্দ্রিক এবং মধ্যমাকেন্দ্রিক গড়বিচ্যুতি নির্ণয় করা যাক।

ক এবং খ-এর রানসংখ্যা যথাক্রমে $x_1$ এবং $x_2$ দ্বারা সূচিত করা হলে,

$$\bar{x}_1 = 60 \qquad \bar{x}_2 = 60$$

$$\tilde{x}_1 = \frac{61+61}{2} = 61 \qquad \approx \; 42+68 \; = 55$$

**সারণী 5.1**

**5.1 অনুচ্ছেদে প্রদত্ত রাশিতথ্যের গড়বিচ্যুতি নির্ণয়**

| ক-এর রান $x_{1i}$ | খ-এর রান $x_{2i}$ | $\lvert x_{1i}-60\rvert$ | $\lvert x_{2i}-60\rvert$ | $\lvert x_{1i}-61\rvert$ | $\lvert x_{2i}-55\rvert$ |
|---|---|---|---|---|---|
| 62 | 119 | 2 | 59 | 1 | 64 |
| 64 | 23 | 4 | 37 | 3 | 32 |
| 63 | 79 | 3 | 19 | 2 | 24 |
| 55 | 0 | 5 | 60 | 6 | 55 |
| 61 | 68 | 1 | 8 | 0 | 13 |
| 56 | 0 | 4 | 60 | 5 | 55 |
| 58 | 42 | 2 | 18 | 3 | 13 |
| 62 | 85 | 2 | 25 | 1 | 30 |
| 61 | 19 | 1 | 41 | 0 | 36 |
| 58 | 165 | 2 | 105 | 3 | 110 |
| মোট 600 | 600 | 26 | 432 | 24 | 432 |

সুতরাং ক ও খ-এর মানসংখ্যার গড়কেন্দ্রিক গড়বিচ্যুতি যথাক্রমে

$$26/10 = 2{\cdot}6 \text{ এবং } 432/10 = 43{\cdot}2$$

এবং মধ্যমাকেন্দ্রিক গড়বিচ্যুতিগুলি যথাক্রমে 2·4 এবং 43·2.

এখানে লক্ষ্য কর, ক-এর ক্ষেত্রে মধ্যমাকেন্দ্রিক গড়বিচ্যুতি অপেক্ষা গড়-কেন্দ্রিক গড়বিচ্যুতির মান বেশী, এবং খ-এর ক্ষেত্রে এ দুটির মান সমান। অর্থাৎ কোন ক্ষেত্রেই প্রথমোক্তটির মান শেষোক্তটি অপেক্ষা বেশী নয়। খ-এর ক্ষেত্রে দুটির মান সমান হওয়ার কারণ হ'ল, মানসংখ্যা যুগ্ম হওয়ায় খ-এর প্রকৃতপক্ষে দুটি মধ্যমা—42 এবং 68, এবং এই দুটি মানের মধ্যবর্তী যে কোন মানের সম্পর্কে গড়বিচ্যুতির পরিমাণ একই হবে। এক্ষেত্রে গড়মান 60 মধ্যমা মান-দুটির অন্তর্বর্তী হওয়ায় মান-দুটি সমান হয়েছে। পক্ষান্তরে, ক-এর ক্ষেত্রে গড় (60) মধ্যমা মান-দুটির (61, 61) বহির্ভূত হওয়ায় প্রথমোক্ত বিচ্যুতির মান অন্যটি অপেক্ষা কম হয়েছে।

**উদা. 5.3** 3.7 সারণীতে প্রদত্ত রাশিতথ্যের মধ্যমাকেন্দ্রিক গড়বিচ্যুতি নিম্নলিখিতভাবে নির্ণয় করা যেতে পারে।

এখানে মধ্যমা 95·45 ডিঃ ফাঃ।

**সারণী 5.2**

**3.7 সারণীতে প্রদত্ত রাশিতথ্যের মধ্যমাকেন্দ্রিক গড়বিচ্যুতি নির্ণয়**

| $x_i$ | $f_i$ | $\lvert x_i - 95{\cdot}45 \rvert$ | $f_i \times (3)$ |
|---|---|---|---|
| (1) | (2) | (3) | (4) |
| 83·45 | 1 | 12 | 12 |
| 86·45 | 7 | 9 | 63 |
| 89·45 | 19 | 6 | 114 |
| 92·45 | 31 | 3 | 93 |
| 95·45 | 37 | 0 | 0 |
| 98·45 | 26 | 3 | 78 |
| 101·45 | 14 | 6 | 84 |
| 104·45 | 14 | 9 | 126 |
| 107·45 | 4 | 12 | 48 |
| মোট | 153 | – | 618 |

স্বতরাং মধ্যমাকেন্দ্রিক গড়বিচ্যুতি $= \frac{618}{153} = 4{\cdot}0392$ ডিঃ ফাঃ।

## 5.5 প্রমাণ বিচ্যুতি (standard deviation) :

**5.5.1 সংজ্ঞা :** চতুর্থ পরিচ্ছেদে বর্ণিত গাণিতিক গড়কে সরল গড় বলা যেতে পারে। পক্ষান্তরে প্রদত্ত কয়েকটি মানের **বর্গগড়** (mean square) হ'ল ঐ মানগুলির বর্গসমূহের গড়।

বিস্তৃতি-মাপকের সূত্রে চিহ্নযুক্ত বিচ্যুতি নেওয়া হলে যে অসুবিধার কথা আগে বলা হ'ল, সেটি বিকল্পভাবে দূর করা যেতে পারে বিচ্যুতিগুলির বর্গগড় নিয়ে। এই জাতীয় একটি বিস্তৃতি-মাপক হ'ল **মুল-গড়-বর্গবিচ্যুতি**। নির্বাচিত মধ্যগামিতা-মাপকটি $A$ দ্বারা নির্দেশিত হলে, $x_1, x_2, \ldots, x_n$—এই $n$টি মানের $A$-কেন্দ্রিক মূল-গড়-বর্গবিচ্যুতি হ'ল

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - A)^2} \qquad \cdots \qquad (5.4)$$

এখানে বর্গমূলটির ধনাত্মক মানটিই নেওয়া হয়।

বিচ্যুতিগুলির বর্গগড়ের পরিবর্তে এর বর্গমূলটি বিস্তৃতি-মাপক হিসাবে নেওয়ার কারণ হ'ল, বর্গগড় প্রদত্ত হয় সংশ্লিষ্ট বর্গ-এককে, কিন্তু বিস্তৃতি-মাপকের একক এবং সংশ্লিষ্ট মানগুলির একক অভিন্ন হওয়া উচিত।

মূল-গড়-বর্গবিচ্যুতি নির্ণয় করতে সাধারণতঃ $\bar{x}$-কে কেন্দ্র হিসাবে ব্যবহার করা হয়। $\bar{x}$-কেন্দ্রিক মূল-গড়-বর্গবিচ্যুতিকে বলা হয় **প্রমাণ বিচ্যুতি**। অর্থাৎ, আলোচ্য ক্ষেত্রে প্রমাণবিচ্যুতি $s$ হ'ল,

$$s=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}. \qquad \cdots \qquad (5.5)$$

শ্রেণীবিন্যস্ত রাশিতথ্যের ক্ষেত্রে প্রমাণবিচ্যুতির সূত্র

$$s=\sqrt{\frac{1}{n}\sum_{i=1}^{k}f_i(x_i-\bar{x})^2}. \qquad \cdots \qquad (5.6)$$

$$\text{এখন} \quad \sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$=\sum_{i=1}^{n}x_i^2+n\bar{x}^2-2\bar{x}\sum_{i=1}^{n}x_i$$

$$=\sum_{i=1}^{n}x_i^2-n\bar{x}^2,$$

$$\text{অর্থাৎ} \quad s=\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2-\bar{x}^2}. \qquad \cdots \qquad (5.5a)$$

অনুরূপভাবে (5.6) এর সরলীকৃত রূপ

$$s=\sqrt{\frac{1}{n}\sum_{i=1}^{k}f_ix_i^2-\bar{x}^2}. \qquad (5.6a)$$

লক্ষণীয়, প্রমাণবিচ্যুতি নির্ণয়ের ক্ষেত্রে (5.5*a*) এবং (5.6*a*) সূত্র দুটি ব্যবহার করাই সুবিধাজনক।

প্রমাণবিচ্যুতির বর্গকে বলা হয় **ভেদমান** (variance)।

**উদা. 5.4** 5.1 অনুচ্ছেদে প্রদত্ত ক ও খ-এর রানসংখ্যার প্রমাণবিচ্যুতি যথাক্রমে

$$s_1 = \left\{\frac{1}{10}\sum_{i=1}^{10} x_{1i}^2 - \bar{x}_1^2\right\}^{\frac{1}{2}} = \sqrt{3608\cdot4 - 3600}$$

$$= \sqrt{8\cdot4} = 2\cdot8983$$

এবং
$$s_2 = \left\{\frac{1}{10}\sum_{i=1}^{10} x_{2i}^2 - \bar{x}_2^2\right\}^{\frac{1}{2}} = \sqrt{6213\cdot5 - 3600}$$

$$= \sqrt{2613\cdot00} = 50\cdot1175.$$

**5.5.2 প্রমাণবিচ্যুতির ধর্মাবলী :** প্রমাণবিচ্যুতির কয়েকটি গাণিতিক ধর্ম আছে ; যথা :—

(i) $X$ চলটির প্রদত্ত মানগুলি ধ্রুবক হলে $X$-এর প্রমাণবিচ্যুতির মান শূন্য।

**প্রমাণ :** মনে কর, $x_i = a,\ i = 1\ (1)\ n.$

তাহলে $\bar{x} = a.$

স্তরাং,
$$s_x = \sqrt{\frac{1}{n}\sum_{1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{1}{n}\sum_{1}^{n}(a-a)^2}$$

$$= 0.$$

(ii) যদি $Y = a + bX$ হয়, তাহলে

$$s_y = |b|\, s_x. \qquad \cdots \qquad (5.7)$$

**প্রমাণ :** $y_i = a + bx_i,\ i = 1\ (1)\ n.$

স্তরাং, $\bar{y} = a + b\bar{x}$

অর্থাৎ,
$$s_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n}\sum_{i=1}^{n}(a + bx_i - a - b\bar{x})^2$$

$$= b^2 . \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = b^2 s_x^2$$

ফলে, $s_y = |b|\, s_x$, কারণ ভেদমানের ধনাত্মক বর্গমূলই প্রমাণবিচ্যুতি।

প্রমাণবিচ্যুতির আলোচ্য ধর্মটি বিভিন্ন পরিস্থিতিতে স্বল্পায়াসে প্রমাণবিচ্যুতি নির্ণয়ে কি-ভাবে সহায়তা করে, নীচের উদাহরণ দুটি থেকে সেটি পরিষ্কার হবে।

**উদা. 5.5** 4.3 সারণীতে প্রদত্ত রাশিতথ্যের প্রমাণবিচ্যুতি নির্ণয়ের কৌশল নীচের সারণীতে প্রদত্ত হল।

### সারণী 5.3

### 4.3 সারণীর অন্তর্গত রাশিতথ্যের প্রমাণবিচ্যুতি নির্ণয়

| $x_i$ | $y_i = x_i - 3000$ | $y_i^2$ |
|---|---|---|
| (1) | (2) | (3) |
| 3125 | 125 | 15625 |
| 3250 | 250 | 62500 |
| 2960 | − 40 | 1600 |
| 3055 | 50 | 2500 |
| 3200 | 200 | 40000 |
| 3125 | 125 | 15625 |
| 2775 | − 225 | 50625 |
| — | 560 | 188475 |

$\bar{y} = 70$ গ্রা., [ উদা. 4.4 ]

এবং $s_y^2 = \frac{1}{7} \times 188475 - 70^2 = 26925 - 4900 = 22025 = s_x^2$.

সুতরাং $s_x = \sqrt{22025} = 148{\cdot}01$ গ্রা.।

লক্ষ্য কর, এইভাবে সংশ্লিষ্ট চলটির রূপান্তর সাধনের ফলে 3125, 3250, প্রভৃতি বড় বড় মানগুলির বর্গনির্ণয়ের পরিশ্রমের অনেকটা লাঘব হয়েছে।

শ্রেণীবিন্যস্ত রাশিতথ্যের ক্ষেত্রে বিভিন্ন শ্রেণীগুলি সমদৈর্ঘ্য হলে এই শ্রম আরও কিছুটা লাঘব করা যায়। নীচের উদাহরণটি লক্ষ্য কর।

**উদা. 5.6.** 3.9 সারণীতে প্রদত্ত রাশিতথ্যের প্রমাণবিচ্যুতি নির্ণয় করা যাক।

সারণী 5.4

3.9 সারণীতে প্রদত্ত রাশিতথ্যের প্রমাণবিচ্যুতি নির্ণয়

| তাপমাত্রা ( শ্রেণী-মধ্যক ( ডিঃ ফাঃ ) | পরিসংখ্যা | $y_i = \frac{x_i - 95{\cdot}45}{3}$ | $y_i f_i$ | $y_i^2 f_i$ |
|---|---|---|---|---|
| 83·45 | 1 | −4 | −4 | 16 |
| 86·45 | 7 | −3 | −21 | 63 |
| 89·45 | 19 | −2 | −38 | 76 |
| 92·45 | 31 | −1 | −31 | 31 |
| 95·45 | 31 | 0 | 0 | 0 |
| 98·45 | 26 | 1 | 26 | 26 |
| 101·45 | 14 | 2 | 28 | 56 |
| 104·45 | 14 | 3 | 42 | 126 |
| 107·45 | 4 | 4 | 16 | 64 |
| মোট | 153 | — | 18 | 458 |

এখানে $s_y = \left[\frac{1}{153}\left\{458 - \frac{(18)^2}{153}\right\}\right]^{\frac{1}{2}}$

$= \sqrt{2{\cdot}9797} = 1{\cdot}7262$ ডিঃ ফাঃ ;

সুতরাং $s_x = 3s_y = 5{\cdot}1786$ ডিঃ ফাঃ ।

(iii) বিভিন্ন গোষ্ঠীর প্রমাণবিচ্যুতি থেকে সার্বিক প্রমাণবিচ্যুতি নির্ণয়ন। মনে কর, একই জাতীয় দুটি বিভিন্ন গোষ্ঠীর প্রথমটিতে আছে $n_1$টি মান, এবং দ্বিতীয়টিতে আছে $n_2$টি। গোষ্ঠী-দুটির গড় এবং প্রমাণবিচ্যুতি যথাক্রমে $\bar{x}_1$ এবং $s_1$ ও $\bar{x}_2$ এবং $s_2$.

ধরা যাক, প্রথম গোষ্ঠীর বিভিন্ন মানগুলি $x_{11}, x_{12}, \ldots, x_{1n_1}$ এবং দ্বিতীয় গোষ্ঠীর বিভিন্ন মানগুলি $x_{21}, x_{22}, \ldots, x_{2n_2}$.

সুতরাং $\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}, \quad i = 1, 2.$

এবং $s_i^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2, \quad i = 1, 2.$

এখন, $n_1+n_2$টি মানের সার্বিক গড় $\bar{\bar{x}}$ এবং প্রমাণবিচ্যুতি $s$ দ্বারা নির্দেশ করা হলে,

স্পষ্টতঃই, $\bar{\bar{x}}=\frac{1}{n}(n_1\bar{x}_1+n_2\bar{x}_2)$, যেখানে $n=n_1+n_2$,

এবং সংজ্ঞানুসারে,

$$s^2=\frac{1}{n}\left[\sum_{j=1}^{n_1}(x_{1j}-\bar{\bar{x}})^2+\sum_{j=1}^{n_2}(x_{2j}-\bar{\bar{x}})^2\right]. \quad \cdots \quad (5.8)$$

$$\sum_{j=1}^{n_i}(x_{ij}-\bar{\bar{x}})^2=\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i+\bar{x}_i-\bar{\bar{x}})^2$$

$$=\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2+n_i(\bar{x}_i-\bar{\bar{x}})^2,$$

$\left[\text{যেহেতু } \sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)=0\right]$

$$=n_is_i^2+n_id_i^2, \quad \text{যেখানে } d_i=\bar{x}_i-\bar{\bar{x}},\ i=1,\ 2.$$

সুতরাং $$s^2=\frac{n_1s_1^2+n_2s_2^2+n_1d_1^2+n_2d_2^2}{n_1+n_2}. \quad \cdots \quad (5.9)$$

(5.9) সূত্রটি সরাসরি সম্প্রসারিত করা যায়। যদি $k$ টি গোষ্ঠী থাকে এবং $i$-তম গোষ্ঠীর মানসংখ্যা, গড় এবং প্রমাণবিচ্যুতি যথাক্রমে $n_i$, $\bar{x}_i$, এবং $s_i$ হয় তাহলে $n=\sum_{i=1}^{k}n_i$ টি মানের সার্বিক প্রমাণবিচ্যুতি $s$ পাওয়া যায় নীচের সূত্র থেকে,

$$s^2=\frac{\sum_{i=1}^{k}n_i(s_i^2+d_i^2)}{\sum_{i=1}^{k}n_i}, \quad \cdots \quad (5.10)$$

যেখানে $d_i = (\bar{x}_i - \bar{\bar{x}})$,

$$\text{এবং} \quad \bar{\bar{x}} = \frac{\sum_{i=1}^{k} n_i \bar{x}_i}{\sum_{i=1}^{k} n_i}.$$

## 5.6 গড় পার্থক্য (mean difference) :

গড়বিচ্যুতি এবং প্রমাণবিচ্যুতি নির্ণয় করা হয় কোন না কোন মধ্যগামিতা-মাপক থেকে। এখন প্রশ্ন হ'ল, কোন চলের বিস্তৃতি পরিমাপের সময় যথেচ্ছ-গৃহীত একটি মধ্যগামিতা-মাপকের ব্যবহার কি অপরিহার্য? প্রদত্ত মানগুলির পারস্পরিক পার্থক্য কতটুকু, বিস্তৃতির অর্থ তাই হওয়া উচিত—সুতরাং অনেকের মতে বিস্তৃতি পরিমাপন প্রসঙ্গে মধ্যগামিতা-মাপকের ব্যবহার অবান্তর। আর তাছাড়া উপযুক্ত মধ্যগামিতা-মাপক নির্বাচনের প্রশ্নটি যেহেতু ব্যক্তিনির্ভর, অতএব এই ধরনের মাপক ব্যবহারে একই রাশিতথ্যের বিস্তৃতি সম্বন্ধে বিভিন্ন জনের বিভিন্ন ধারণা হওয়া সম্ভব।

এই অসুবিধা দূর করার উদ্দেশ্যে ইতালীয় রাশিবিজ্ঞানী গিনি (Gini) প্রদত্ত মানগুলির পারস্পরিক পার্থক্যের ভিত্তিতে নিম্নলিখিত বিস্তৃতি-মাপকটি ব্যবহারের সপক্ষে মত প্রকাশ করেছেন। রাশিবিজ্ঞানের পরিভাষায় এটিকে **গড় পার্থক্য** বলা হয়। গড় পার্থক্য $\Delta_1$-এর সূত্র হ'ল :

$$\Delta_1 = \begin{cases} \dfrac{1}{n^2} \displaystyle\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|, & \cdots \quad (5.11) \\ \quad \text{অবিন্যস্ত রাশিতথ্যের ক্ষেত্রে} & \\ \dfrac{1}{n^2} \displaystyle\sum_{i=1}^{k} \sum_{j=1}^{k} f_i f_j |x_i - x_j|, & \cdots \quad (5.12) \\ \quad \text{শ্রেণীবিন্যস্ত রাশিতথ্যের ক্ষেত্রে।} & \end{cases}$$

$\Delta_1$-এর সূত্রটিতে ভাজক হিসাবে $n^2$ নেওয়ার কারণ হ'ল, এখানে $|x_i - x_j|$ এই জাতীয় পার্থক্যের মোট সংখ্যা $n^2$. পার্থক্যের পরমমান নেওয়ার পরিবর্তে আগের মত বর্গগড় নেওয়া যেতে পারে। সেক্ষেত্রে গড় পার্থক্যের পরিবর্তিত দ্বিতীয় রূপ দাঁড়াবে,

$$\Delta_2 = \begin{cases} \sqrt{\left\{\frac{1}{n^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\left(x_i - x_j\right)^2\right\}}, & \cdots (5.13) \\ & \text{অবিন্যস্ত রাশিতথ্যের ক্ষেত্রে} \\ \sqrt{\left\{\frac{1}{n^2}\sum_{i=1}^{k}\sum_{j=1}^{k} f_i f_j\left(x_i - x_j\right)^2\right\}}, & (5.14) \\ & \text{শ্রেণীবিন্যস্ত রাশিতথ্যের ক্ষেত্রে।} \end{cases}$$

5.7 অনুচ্ছেদে দেখা যাবে, $\Delta_2$ এবং $s$-এর মধ্যে একটি সরাসরি সম্পর্ক বিদ্যমান। অর্থাৎ, প্রমাণবিচ্যুতি জানা থাকলে $\Delta_2$ রাশিতথ্যের বিস্তৃতির ওপর অতিরিক্ত কোন আলোকপাত করতে সক্ষম হয় না।

## 5.7 বিভিন্ন বিস্তৃতি-মাপক সংক্রান্ত কয়েকটি ফল :

(i) মধ্যমাকেন্দ্রিক গড়বিচ্যুতির পরিমাণ লঘিষ্ঠ।

**প্রমাণ :** $A$ যদি একটি যথেচ্ছ-গৃহীত মান হয়, আমাদের প্রমাণ করতে হবে যে,

$$\frac{1}{n}\sum_{i=1}^{n}|x_i - \tilde{x}| < \frac{1}{n}\sum_{i=1}^{n}|x_i - A|. \qquad \cdots (5.15)$$

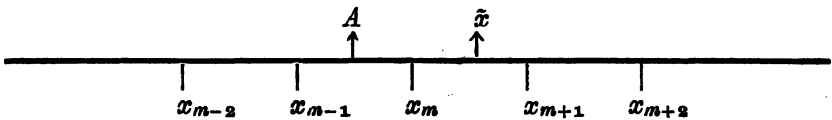
মনে কর, $x_1 < x_2 < \cdots\cdots < x_n$.

প্রথমে ধরা যাক, $n = 2m$.

সুতরাং $x_m \leqslant \tilde{x} \leqslant x_{m+1}$.

এখন মনে কর, $A < \tilde{x}$. এক্ষেত্রে যদি $x_{m-1} \leqslant A \leqslant x_m$ হয়, তাহলে $d = \tilde{x} - A$ লিখে পাই,

$A$ ↑ $\tilde{x}$ ↑

$x_{m-2}$ $x_{m-1}$ $x_m$ $x_{m+1}$ $x_{m+2}$

$$\begin{aligned}\Delta &= \sum_{i=1}^{n}|x_i - A| - \sum_{i=1}^{n}|x_i - \tilde{x}| \\ &= -(m-1)d + md + (x_m - A) - (\tilde{x} - x_m) \\ &= d - \{(\tilde{x} - x_m) - (x_m - A)\} \geqslant d - \{(\tilde{x} - x_m) + (x_m - A)\} \\ &= 0.\end{aligned}$$

আবার, $x_{m-2} < A < x_{m-1}$ হলে, $d' = \tilde{x} - A$ লিখে পাওয়া যায়,

$$\Delta = -(m-2)\,d' + md' - \{(\tilde{x} - x_m) - (x_m - A)\}$$
$$- \{(\tilde{x} - x_{m-1}) - (x_{m-1} - A)\}$$
$$> 2d' - \{(\tilde{x} - x_m) + (x_m - A)\} - \{(\tilde{x} - x_{m-1}) + (x_{m-1} - A)\} = 0.$$

এইভাবে দেখানো যেতে পারে, $A < \tilde{x}$ হলে $A$ এর যে কোন অবস্থিতির জন্য $\Delta > 0$.

এবার মনে কর, $A > \tilde{x}$. এক্ষেত্রে $x_{m+1} < A < x_{m+2}$ হলে, $d_1 = A - \tilde{x}$ লিখে পাওয়া যায়

$$\Delta = md_1 - (m-1)d_1 + (A - x_{m+1}) - (x_{m+1} - \tilde{x})$$
$$= d_1 - \{(x_{m+1} - \tilde{x}) - (A - x_{m+1})\}$$
$$> d_1 - \{(x_{m+1} - \tilde{x}) + (A - x_{m+1})\} = 0.$$

তেমনি, $x_{m+2} < A < x_{m+3}$, $x_{m+3} < A < x_{m+4}$, ...... প্রভৃতি ক্ষেত্রেও $\Delta > 0$.

লক্ষ্য কর, $n = 2m$-এর ক্ষেত্রে $x_m < A < \tilde{x}$ বা $\tilde{x} < A < x_{m+1}$ হলে $\Delta = 0$.

সুতরাং $n = 2m$ হলে (5.15) সম্পর্কটি সত্য।

অনুরূপভাবে দেখানো যেতে পারে $n = 2m + 1$ হলেও সম্পর্কটি সত্য।

(ii) প্রমাণবিচ্যুতিই লঘিষ্ঠ মূল-গড়-বর্গবিচ্যুতি।

**প্রমাণ ঃ** $$\sum_{i=1}^{n} (x_i - A)^2 = \sum_{i=1}^{n} \left\{(x_i - \bar{x}) + (\bar{x} - A)\right\}^2$$

$$= \sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - A)^2, \text{ যেহেতু } \sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

$$> \sum_{i=1}^{n} (x_i - \bar{x})^2, \text{ যেহেতু } (\bar{x} - A)^2 > 0,\ n > 1.$$

সুতরাং $$\sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - A)^2} > \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

( প্রমাণিত )। ··· (5.16)

(iii) গড়কেন্দ্রিক গড়বিচ্যুতির মান প্রমাণবিচ্যুতি অপেক্ষা বৃহত্তর হতে পারে না।

**প্রমাণ :** মনে কর, শ্রেণীবিন্যস্ত রাশিতথ্যে $k$ সংখ্যক শ্রেণী আছে। $i$-তম শ্রেণীর মধ্যক $x_i$ এবং পরিসংখ্যা $f_i$, $i=1\,(1)\,k$. সুতরাং আমাদের প্রতিপাদ্য বিষয়

$$\frac{1}{n}\sum_{i=1}^{k} f_i\,|x_i - x| < \sqrt{\frac{1}{n}\sum_{i=1}^{k} f_i\,(x_i-\bar{x})^2} \qquad (5.17)$$

এখন **কোশি-শোয়াৎর্জের অসমতা** (Cauchy Scwartz Inequality) সম্পর্কটি হল, $u_1, u_2, \cdots u_n$ এবং $v_1, v_2, \cdots, v_n$ যদি দুই প্রস্থ বাস্তব (real) রাশি হয়, তাহলে,

$$\left(\sum_{i=1}^{k} u_i\,v_i\right)^2 < \left(\sum_{i=1}^{k} u_i^2\right)\left(\sum_{i=1}^{k} v_i^2\right). \qquad \cdots \quad (5.18)$$

এই অসমতা-সম্পর্কটি আগে প্রমাণ করা যাক।

স্পষ্টতঃই, $(|a_i| - |b_i|)^2 \geqslant 0, \quad i=1\,(1)\,n.$

ফলে, $a_1^2 + b_1^2 \geqslant 2|a_1|\,|b_1|$

$a_2^2 + b_2^2 \geqslant 2|a_2|\,|b_2|$

$a_n^2 + b_n^2 \geqslant 2|a_n|\,|b_n|$

সুতরাং $$\sum_i a_i^2 + \sum_i b_i^2 \geqslant 2\sum_i |a_i|\,|b_i| \geqslant 2\left|\sum_i a_i b_i\right.$$

এখন মনে কর, $$a_i = \frac{u_i}{\sqrt{\sum_i u_i^2}}$$

এবং $$b_i = \frac{v_i}{\sqrt{\sum_i v_i^2}},\; i=1(1)n$$

সুতরাং, $$\sum_i \left\{\frac{u_i^2}{\sum_i u_i^2}\right\} + \sum_i \left\{\frac{v_i^2}{\sum_i v_i^2}\right\}$$

$$\geqslant 2 \frac{\left|\sum u_i v_i\right|}{\sqrt{\sum_i u_i^2}\sqrt{\sum_i v_i^2}}$$

অর্থাৎ, $$2 \geqslant 2 \frac{\left|\sum u_i v_i\right|}{\sqrt{\sum_i u_i^2}\sqrt{\sum_i v_i^2}}$$

অর্থাৎ, $$\sqrt{\sum_i u_i^2}\sqrt{\sum_i v_i^2} \geqslant \left|\sum_i u_i v_i\right|$$

অর্থাৎ, $$\sum_i u_i^2 \sum_i v_i^2 \geqslant \left(\sum_i u_i v_i\right)^2 \text{......( প্রমাণিত )।}$$

এখন এই সূত্রে $u_i = \sqrt{f_i}\,|x_i - \bar{x}|$ এবং $v_i = \sqrt{f_i}$ বসিয়ে পাওয়া যায়

$$\left(\sum_i f_i\,|x_i - \bar{x}|\right)^2 \leqslant \sum_i f_i (x_i - \bar{x})^2 . n$$

অর্থাৎ, $$\frac{1}{n}\sum_i f_i |x_i - \bar{x}| \leqslant \sqrt{\frac{1}{n}\sum f_i (x_i - \bar{x})^2} \text{ ...( প্রমাণিত )।}$$

**টীকা।** প্রমাণবিচ্যুতি, গড়বিচ্যুতি এবং আন্তঃচতুর্থক অর্ধপ্রসারের মধ্যে নিম্নলিখিত অবেক্ষণভিত্তিক (empirical) সম্পর্ক-দুটি বাস্তবে দৃষ্ট অনেক বিভাজনের ক্ষেত্রে সত্য :

(i) গড়বিচ্যুতি $\simeq \frac{4}{5}$ প্রমাণবিচ্যুতি, ... (5.19)

এবং, (ii) আন্তঃচতুর্থক অর্ধপ্রসার $\simeq \frac{2}{3}$ প্রমাণবিচ্যুতি। ... (5.20)

(iv) $x_1, x_2, \ldots, x_n$—এই $n$টি মানের প্রসার এবং প্রমাণবিচ্যুতি যথাক্রমে $R$ এবং $s$ দ্বারা নির্দেশ করা হলে,

$$\frac{R^2}{2n} \leqslant s^2 \leqslant \frac{R^2}{4}. \qquad \cdots \qquad \cdots \quad (5.21)$$

**প্রমাণ :** ধরা যাক, $x_1 \leqslant x_2 \leqslant \cdots\cdots \leqslant x_n$.

তাহলে $R = x_n - x_1$.

এখন, $$ns^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$\leqslant \sum_{i=1}^{n} \left(x_i - \frac{x_1 + x_n}{2}\right)^2 \qquad \text{[ অনুচ্ছেদ 5.7(ii) ]}$$

$$= \sum_{x_i > \frac{x_1 + x_n}{2}} \left(x_i - \frac{x_1 + x_n}{2}\right)^2 + \sum_{x_i \leqslant \frac{x_1 + x_n}{2}} \left(x_i - \frac{x_1 + x_n}{2}\right)^2$$

$$\leqslant \sum_{x_i > \frac{x_1 + x_n}{2}} \left(x_n - \frac{x_1 + x_n}{2}\right)^2 + \sum_{x_i \leqslant \frac{x_1 + x_n}{2}} \left(x_1 - \frac{x_1 + x_n}{2}\right)^2$$

$$= \sum_{x_i > \frac{x_1 + x_n}{2}} \frac{R^2}{4} + \sum_{x_i \leqslant \frac{x_1 + x_n}{2}} \frac{R^2}{4}$$

$$= \sum_{i=1}^{n} \frac{R^2}{4} = \frac{nR^2}{4}, \text{ অর্থাৎ, } s^2 \leqslant \frac{R^2}{4}. \qquad \cdots \quad (5.21a)$$

আবার, $$ns^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$\geqslant (x_1 - \bar{x})^2 + (x_n - \bar{x})^2$$

$$\geqslant \left(x_1 - \frac{x_1 + x_n}{2}\right)^2 + \left(x_n - \frac{x_1 + x_n}{2}\right)^2 \qquad \text{[ অনুচ্ছেদ 5.7(ii) ]}$$

$$= \frac{R^2}{4} + \frac{R^2}{4} = \frac{R^2}{2}$$

বা, $s^2 \geqslant R^2/2n$. $\qquad \cdots \qquad \cdots \quad (5.21b)$

সুতরাং, (5.21$a$) এবং (5.21$b$) থেকে,

$$\frac{R^2}{2n} \leqslant s^2 \leqslant \frac{R^2}{4} \qquad \cdots \quad \text{( প্রমাণিত )।}$$

(v) $\Delta_2{}^2 = 2s^2$. ... ... ... (5.22)

**প্রমাণ :** $n^2\Delta_2{}^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - x_j)^2$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} [(x_i - \bar{x}) - (x_j - \bar{x})]^2$$

$$= n \sum_{i=1}^{n} (x_i - \bar{x})^2 + n \sum_{j=1}^{n} (x_j - \bar{x})^2, \quad \because \sum_{i=1}^{n} (x_i - \bar{x}) = 0;$$

$$= n^2s^2 + n^2s^2$$

ফলে, $\Delta_2{}^2 = 2s^2$.........( প্রমাণিত )।

**5.8 আপেক্ষিক বিস্তৃতি-মাপক** (relative measures of dispersion) :

কোন চলের নির্বাচিত একটি বিস্তৃতি-মাপকের মানকে চলটির নির্বাচিত একটি মধ্যগামিতা-মাপকের মানের অনুপাতে ( বা শতকরা পরিমাণে ) প্রকাশ করা হলে পাওয়া যায় চলটির **বিস্তৃতি-অঙ্ক** (coefficient of dispersion)। প্রকৃতপক্ষে চলটির আপেক্ষিক বিস্তৃতি মাপনার উদ্দেশ্যে এই মাপকটি ব্যবহৃত হয়।

সর্বাপেক্ষা বহুল ব্যবহৃত বিস্তৃতি-অঙ্ক হ'ল **ভেদাঙ্ক** (coefficient of variation)। স্বাভাবিক অর্থবাহী সঙ্কেতচিহ্নের ব্যবহারে ভেদাঙ্ক $v$-এর সূত্র হ'ল

$$v = \frac{s}{\bar{x}} \times 100. \quad \cdots \quad (5.23)$$

( এখানে $\bar{x}$-এর মান শূন্যেতর ধ'রে নেওয়া হয়েছে )

**চতুর্থক বিচ্যুতি-অঙ্ক** (coefficient of quartile deviation) হ'ল আর একটি আপেক্ষিক বিস্তৃতি মাপক। এর সূত্র

$$\text{চতুর্থক বিচ্যুতি-অঙ্ক} = \frac{2Q}{Q_3 + Q_1} \times 100 = \frac{Q_3 - Q_1}{Q_3 + Q_1} \times 100. \quad \cdots \quad (5.24)$$

( এখানেও $Q_3 + Q_1$-এর মান শূন্যেতর ধ'রে নেওয়া হয়েছে )

শ্রেণীবিন্যস্ত রাশিতথ্যের ক্ষেত্রে এক বা একাধিক প্রান্তশ্রেণী অনির্দিষ্ট দৈর্ঘ্যসম্পন্ন হলে $s$ বা $\bar{x}$-এর যথার্থ মান নির্ণয় করা সম্ভব হয় না। সেক্ষেত্রে আপেক্ষিক বিস্তৃতি-মাপক হিসাবে চতুর্থক বিচ্যুতি-অঙ্কের ব্যবহার অপরিহার্য।

অনেক সময় বিস্তৃতির বিচারে একাধিক চলের তুলনা করার প্রয়োজন হয়। সেক্ষেত্রে দু'ধরনের অসুবিধা দেখা দিতে পারে। প্রথমতঃ, চল-দুটির মাপনা-একক

ভিন্ন হতে পারে, সেক্ষেত্রে অনাপেক্ষিক বিস্তৃতি-মাপকের মানগুলির তুলনা অবান্তর। যেমন মনে কর, 100 জন ছাত্রের ওজনের ও উচ্চতার প্রমাণবিচ্যুতি যথাক্রমে 5·6 কিলোগ্রাম এবং 0·65 মিটার—এক্ষেত্রে 5·6 কিলোগ্রাম এবং 0·65 মিটার এই রাশি-দুটির তুলনা হাস্যকর। আবার, এমন হতে পারে, চল-দুটির মাপনা-একক অভিন্ন হলেও এদের মধ্যগামিতার পার্থক্য বিরাট। সেক্ষেত্রে একই পরিমাণ বিস্তৃতি উভয়ক্ষেত্রে এক অর্থবাহী নয়। যেমন মনে কর, দু'ধরনের প্যাকেটে চিনি আছে। বড় মাপের 100টি প্যাকেটে চিনির গড় ওজন 100 কি.গ্রা., প্রমাণবিচ্যুতি 5·5 কি.গ্রা.। ছোট মাপের অন্য 100টি প্যাকেটের চিনির গড় ওজন 250 গ্রাম, প্রমাণবিচ্যুতি 10 গ্রাম। স্পষ্টতঃই প্রথম পরিস্থিতিতে 1 গ্রাম পরিমিত প্রমাণবিচ্যুতি দ্বিতীয় পরিস্থিতিতে 1 গ্রাম পরিমিত প্রমাণবিচ্যুতির সমতুল নয়। সুতরাং দেখা যাচ্ছে, বিস্তৃতির পরিমাণগত প্রকৃতি চলের মধ্যগামিতা বা অবস্থিতির উপর একান্ত নির্ভরশীল।

লক্ষ্য কর, আপেক্ষিক বিস্তৃতি-মাপকের ব্যবহারে উপরোক্ত দুটি অসুবিধাই দূর করা যায়।

**উদা. 5.7** 3.9 সারণীতে প্রদত্ত রাশিতথ্যের জন্য

$\bar{x} = 95{\cdot}8028$ ডিঃ ফাঃ

$s = 5{\cdot}1786$ ডিঃ ফাঃ

$Q_1 = 92{\cdot}0387$ ডিঃ ফাঃ

$Q_3 = 99{\cdot}2283$ ডিঃ ফাঃ

সুতরাং ভেদাঙ্ক $= \dfrac{5{\cdot}1786}{95{\cdot}8028} \times 100 = 5{\cdot}4055,$

এবং চতুর্থক বিচ্যুতি-অঙ্ক $= \dfrac{99{\cdot}2283 - 92{\cdot}0387}{99{\cdot}2283 + 92{\cdot}0387} \times 100$

$= 3{\cdot}7589.$

## 5.9 আদর্শ বিস্তৃতি-মাপক হিসাবে প্রসার, গড়-বিচ্যুতি এবং প্রমাণবিচ্যুতির তুলনা :

প্রসার, গড়বিচ্যুতি এবং প্রমাণবিচ্যুতি—এই তিনটি হ'ল অপেক্ষাকৃত বহুল ব্যবহৃত বিস্তৃতি-মাপক। বর্তমান অনুচ্ছেদে এই তিনটি মাপকের তুলনামূলক আলোচনা করা হচ্ছে।

সংজ্ঞার স্পষ্টতার বিচারে তিনটি মাপকই মোটামুটি সমতুল। তবে গোষ্ঠীবদ্ধ

রাশিতথ্যের ক্ষেত্রে প্রান্তিক শ্রেণী-দুটির যে কোনটি অনির্দিষ্ট দৈর্ঘ্যসম্পন্ন হলে প্রমাণবিচ্যুতি বা গড়বিচ্যুতি নির্ণয় করায় অসুবিধা হয়। এক্ষেত্রে চতুর্থক পার্থক্য একমাত্র সন্তোষজনক সমাধান। তত্ত্বগত নিবেশনের ক্ষেত্রে ( অষ্টম পরিচ্ছেদে আলোচিতব্য ) কোনও চলের মানসীমা একদিকে বা উভয়দিকে অসীম হলে প্রসারের মানও হয় অসীম, কিন্তু অন্যদুটির মান সসীম হতে পারে। আবার গড়বিচ্যুতি কিছুটা ব্যক্তিনির্ভর, কেননা গড়বিচ্যুতি নির্ণয়ে সংশ্লিষ্ট মধ্যগামিতা-মাপকের নির্বাচন সকলের ক্ষেত্রে অভিন্ন নাও হতে পারে।

প্রসার এবং গড়পার্থক্যের তাৎপর্য সহজেই অনুধাবনযোগ্য—সেই তুলনায় বরং প্রমাণবিচ্যুতির প্রকৃতি কিছুটা জটিল (abstract)। এখানে মাপনা-একক ঠিক রাখার জন্য প্রথমে বিচ্যুতিগুলির বর্গগড় নিয়ে পরে বর্গমূল নেওয়া হয়।

স্বল্পায়াসে নিরূপণ-যোগ্যতার বিচারে প্রসার নিঃসন্দেহে তিনটির মধ্যে শ্রেষ্ঠত্বের দাবীদার। এই কারণেই যেসব পরিস্থিতিতে যথাসম্ভব অল্প সময়ের মধ্যে সংগৃহীত রাশিতথ্যের বিস্তৃতির পরিমাণ জানা দরকার [ যেমন, **রাশিবিজ্ঞান-সম্মত গুণনিয়ন্ত্রণের** ক্ষেত্রে, ] সেখানে বিস্তৃতি-মাপক হিসাবে প্রসারের প্রচলন খুব বেশী।

বিস্তৃতির সঠিক সংজ্ঞানুসারে প্রসার বিস্তৃতিমাপক হিসাবে অবশ্যই কিছুটা নিকৃষ্টতর। পরোক্ষভাবে প্রদত্ত সবকটি মানের ওপর নির্ভরশীল হলেও, প্রকৃতপক্ষে এটি কেবলমাত্র প্রান্তিক মান-দুটির ওপর নির্ভর করে। প্রান্তিক মান-দুটি অপরিবর্তিত রেখে অন্যান্য মানগুলির পরিবর্তন সাধন করলে বিস্তৃতির তারতম্য ঘটে ঠিকই, কিন্তু প্রসারের মানের কোন পরিবর্তন হয় না। কিন্তু প্রদত্ত সবকটি মানের ওপর প্রত্যক্ষভাবে নির্ভরশীল হওয়ায় প্রমাণবিচ্যুতি এবং গড়বিচ্যুতির ক্ষেত্রে এই ত্রুটিটি অনুপস্থিত। প্রসার যে চলের বিস্তৃতির সঠিক চিত্র দিতে সক্ষম হয় না, একটি উদাহরণ নিলে সেটা সহজে বোঝা যাবে। মনে কর, 8টি বিভিন্ন পত্রে দুটি ছাত্রের প্রাপ্ত নম্বর হ'ল

ক 10, 10, 10, 10, 90, 90, 90, 90

খ 10, 48, 50, 50, 52, 56, 58, 90

এখানে যদিও ক ও খ-এর প্রাপ্ত নম্বরের প্রসার একই অর্থাৎ $90-10=80$, স্পষ্টতঃই ক-এর প্রাপ্ত নম্বরের বিস্তৃতি অনেক বেশী এবং এখানে প্রমাণবিচ্যুতি বা গড়বিচ্যুতির নিরূপিত মানে ব্যাপারটির সঠিক প্রতিফলন ঘটবে।

প্রমাণবিচ্যুতির বিভিন্ন গাণিতিক ধর্মগুলির জন্য পরবর্তী পর্যায়ে বিভিন্ন

গাণিতিক পদ্ধতি প্রয়োগের পক্ষে এটি খুবই উপযোগী। অন্য দুটি মাপকের তুলনায় প্রমাণবিচ্যুতির এটি একটি বড় সুবিধা।

প্রসার অপেক্ষা প্রমাণবিচ্যুতির নমুনাজ বিচ্যুতি সাধারণতঃ কম হয়।

সুবিধা অসুবিধাগুলির আপেক্ষিক গুরুত্বের বিচারে আলোচ্য তিনটি মাপকের মধ্যে প্রমাণবিচ্যুতিকেই শ্রেষ্ঠ বলা চলে।

## 5.10 কেন্দ্রীভবন রেখা (curve of concentration) :

আয়, সম্পত্তি ইত্যাদির বণ্টনসংক্রান্ত সমীক্ষায় প্রায়ই দেখা যায় এই সমস্ত ক্ষেত্রে বণ্টন-বৈষম্য খুবই প্রকট। যেমন, কোন শহরে সমীক্ষা চালালে দেখা যাবে, শহরের অধিবাসীদের যদি আয়ের উর্ধ্বক্রমে সাজানো হয় তাহলে মোট অধিবাসীদের অর্ধাংশের ( যারা নীচের দিকে অবস্থিত ) বরাতে জুটেছে হয়তো মোট আয়ের শতকরা 20 ভাগ, তিন-চতুর্থাংশের বরাতে হয়তো মাত্র 40 ভাগ, ইত্যাদি—অর্থাৎ মোট আয়ের সিংহভাগ রয়েছে উচ্চবিত্তদের দখলে। অর্থাৎ এইসব ক্ষেত্রে সংশ্লিষ্ট চলটি গৃহীত প্রসারের মধ্যে সমভাবে নিবেশিত নয়। এই ধরনের চলের নিবেশনবৈষম্য বা বিশেষ একটি দিকে কেন্দ্রীভবণের প্রবণতা পরিমাপ করার জন্য **কেন্দ্রীভবন রেখা** বা **লরেঞ্জ রেখা** (Lorenz curve) ব্যবহার করা হয়।

প্রকৃতপক্ষে কেন্দ্রীভবন রেখা হ'ল একধরনের ক্রমযৌগিক পরিসংখ্যা রেখা। $X$ চলের $x$ মানটি পর্যন্ত মোট মানসমষ্টির শতকরা অনুপাত ধরা যাক $\Phi(x)$ এবং এই $x$ মানটি পর্যন্ত শতকরা পরিসংখ্যা ধরা যাক $F(x)$. অর্থাৎ,

$$\Phi(x)=100\sum_{x_i\leq x} f_i\,x_i\Big/\sum_i f_i\,x_i \text{ এবং } F(x)=100\sum_{x_i\leq x} f_i\Big/\sum_i f_i$$

স্পষ্টতঃই $F(x)$ এবং $\Phi(x)$ উভয়েরই মানসীমা 0 থেকে 100. $x$-এর কয়েকটি নির্বাচিত মানের জন্য প্রথমে $F(x)$ এবং $\Phi(x)$-এর মান নির্ণয় করা হয়। অনুভূমিক অক্ষরেখায় $F(x)$ এবং উল্লম্ব অক্ষরেখায় $\Phi(x)$ নির্দেশ ক'রে তারপর $\{F(x), \Phi(x)\}$ বিন্দুগুলি সংস্থাপন করা হয়। হস্তাঙ্কিত মসৃন রেখাদ্বারা সন্নিহিত বিন্দুগুলি যোগ ক'রে যে রেখাটি পাওয়া যায় সেইটিই হ'ল কেন্দ্রীভবন রেখা বা লরেঞ্জ রেখা। $\Phi=F$ এই রেখাটিকে বলা হয় **সমনিবেশনী রেখা** (line of equal distribution বা egalitarian line), কারণ বণ্টনব্যবস্থায় কোনরূপ বৈষম্য না থাকলে কেন্দ্রীভবন রেখাটি $\Phi=F$ এই সরলরেখায় পর্যবসিত হয়।

আয়, সম্পত্তি প্রভৃতি চলের ক্ষেত্রে কেন্দ্রীভবন রেখাটি হয় **ওপরের দিকে অবতল** (concave upward)। প্রাপ্ত কেন্দ্রীভবন রেখা এবং সমনিবেশনী রেখার মধ্যবর্তী অঞ্চলটিকে বলা হয় **কেন্দ্রীভবনাঞ্চল** (area of concentration)। স্পষ্টতঃ, নিবেশনবৈষম্য যত বেশী হবে—অর্থাৎ আমাদের বর্তমান উদাহরণে মোট আয়ের সিংহভাগ যত বেশী পরিমাণে ভাগ্যবানদের করায়ত্ত থাকবে—এই কেন্দ্রীভবনাঞ্চলের আয়তনও তত বেশি হবে। কেন্দ্রীভবনাঞ্চলের আয়তনটিকে তাই নিবেশনবৈষম্য বা কেন্দ্রীভবনের মাপক হিসাবে নেওয়া যেতে পারে। বস্তুতঃ গিনির (Gini) **কেন্দ্রীভবনাঙ্ক** (coefficient of concentration) হ'ল এই কেন্দ্রীভবনাঞ্চলের দ্বিগুণিত আয়তন।

**উদা. 5.8** নীচের সারণীতে একটি কারখানার 1,080 জন শ্রমিকের সাপ্তাহিক আয়ের পরিসংখ্যা-বিভাজন দেওয়া হয়েছে।

**সারণী 5.5**

**1,080 জন শ্রমিকের সাপ্তাহিক আয়ের পরিসংখ্যা-বিভাজন**

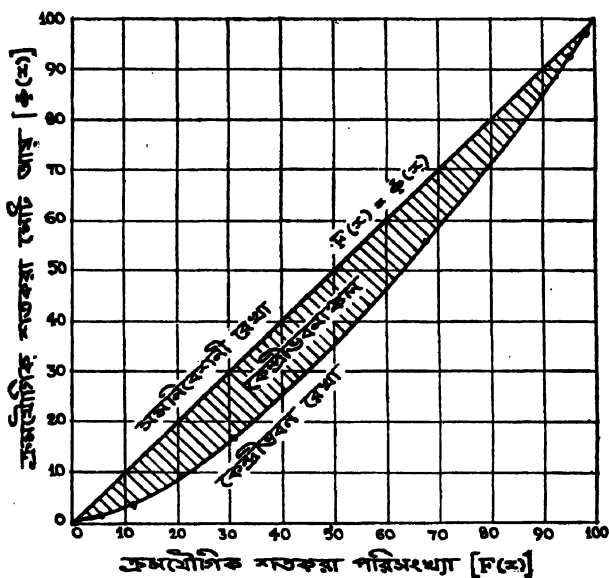
| সাপ্তাহিক আয় ( টাকায় ) | পরিসংখ্যা |
|---|---|
| 1—10 | 50 |
| 10—19 | 70 |
| 19—28 | 203 |
| 28—37 | 406 |
| 37—46 | 304 |
| 46—55 | 42 |
| 55—64 | 5 |
| মোট | 1,080 |

এক্ষেত্রে কেন্দ্রীভবন রেখাটি অঙ্কিত করা যাক। এর জন্য প্রথমে নিম্নলিখিত ছক অনুযায়ী অঙ্কপাতন করতে হবে।

## সারণী 5.6

## 5.5 সারণীতে প্রদত্ত রাশিতথ্যের কেন্দ্রীভবন রেখা অঙ্কনের জন্য প্রয়োজনীয় অঙ্কপাতন

| আয় ( টাকায় ) $x$ ( শ্রেণী-মধ্যক ) | পরিসংখ্যা $f$ | মোট আয় $xf$ | ক্রমযৌগিক মোট আয় | ক্রমযৌগিক পরিসংখ্যা | ক্রমযৌগিক শতকরা পরিসংখ্যা $F(x)$ | ক্রমযৌগিক শতকরা মোট আয় $\Phi(x)$ |
|---|---|---|---|---|---|---|
| 5·5 | 50 | 275·00 | 275·00 | 50 | 4·63 | 0·80 |
| 14·5 | 70 | 1,015·00 | 1,290·00 | 120 | 11·11 | 3·76 |
| 23·5 | 203 | 4,770·50 | 6,060·50 | 323 | 29·91 | 17·67 |
| 32·5 | 406 | 13,195·00 | 19,255·50 | 729 | 67·50 | 56·15 |
| 41·5 | 304 | 12,616·00 | 31,871·50 | 1,033 | 95·65 | 92·95 |
| 50·5 | 42 | 2,121·00 | 33,992·50 | 1,075 | 99·54 | 99·13 |
| 59·5 | 5 | 297·50 | 34,290·00 | 1,080 | 100·00 | 100·00 |
| মোট | 1,080 | 34,290·00 | — | — | — | — |

নীচে 5.1 চিত্রে কেন্দ্রীভবন রেখাটি অঙ্কিত হ'ল। সংখ্যাভিত্তিক গণিতের [ পরিশিষ্ট দ্রষ্টব্য ] কোন আসন্ন পদ্ধতি ব্যবহার ক'রে কেন্দ্রীভবনাঙ্কের মান নির্ণয় করা যাবে।

চিত্র 5.1

5.5 সারণীতে প্রদত্ত রাশিতথ্যের কেন্দ্রীভবন রেখা এবং কেন্দ্রীভবনাঙ্কল।

## 5.11 অনুশীলনী

5.1 বিস্তৃতির সংজ্ঞা দাও। দুটি সমজাতীয় পরিসংখ্যা-বিভাজনের মধ্যে তুলনা করার জন্য মধ্যগামিতা ছাড়াও বিস্তৃতির বিচার করা প্রয়োজন কেন, উদাহরণ সহযোগে ব্যাখ্যা কর। প্রচলিত বিস্তৃতি-মাপকগুলির উল্লেখ কর।

5.2 আদর্শ বিস্তৃতি-মাপকের কী কী ধর্ম থাকা উচিত? এই প্রসঙ্গে বিস্তৃতি-মাপক হিসাবে প্রসার, গড়বিচ্যুতি এবং প্রমাণবিচ্যুতির তুলনা কর।

5.3 আপেক্ষিক বিস্তৃতি-মাপকের প্রয়োজনীয়তা আলোচনা কর। প্রচলিত আপেক্ষিক বিস্তৃতি-মাপকগুলির উল্লেখ কর। প্রমাণ কর, একটি প্রতিসম

বিভাজনের ক্ষেত্রে চলটির মানগুলি অঋণাত্মক হলে ভেদাঙ্কের মান অবশ্যই 0 এবং 100-এর মধ্যে অবস্থান করে।

5.4 প্রমাণবিচ্যুতির সংজ্ঞা দাও। মাপকটির বিভিন্ন গাণিতিক ধর্ম আলোচনা কর। দেখাও যে (5.9) সূত্রটির বিকল্প একটি রূপ

$$s^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2} + \frac{n_1 n_2 (\bar{x}_1 - \bar{x}_2)^2}{(n_1 + n_2)^2}.$$

5.5 যদি $a \leqslant x_i \leqslant b$, $i = 1\,(1)n$ হয়, তাহলে এই $n$টি মানের প্রমাণ-বিচ্যুতি $s$ হলে দেখাও যে

$$0 \leqslant s^2 \leqslant (b-a)^2/4.$$

সম্পর্কটি কখন সমতায় পর্যবসিত হবে? দুটি সংখ্যার প্রমাণবিচ্যুতি $= 6$. এদের একটি 8 হলে অপরটি কত?

[ইঙ্গিত: (5.21) সূত্র দেখ। এখানে $a \leqslant x_{(1)}$ এবং $b \geqslant x_{(n)}$]

5.6 একটি পরিসংখ্যা-বিভাজনে $k$টি সমদৈর্ঘ্য শ্রেণী আছে। $w =$ শ্রেণী-দৈর্ঘ্য, $x_i$ এবং $f_i$ যথাক্রমে $i$-তম শ্রেণীর মধ্যক এবং পরিসংখ্যা,

$$f_i' = \sum_{j=1}^{i} f_j,\; f_i'' = \sum_{j=1}^{i} f_j',\; F_1 = \sum_{i=1}^{k} f_i'/n,\; F_2 = \sum_{i=1}^{k} f_i''/n,$$

এবং $n = \sum_{i=1}^{k} f_i$ হলে প্রমাণ কর,

$$\bar{x} = x_k - w(F_1 - 1),$$

এবং $$s = w\sqrt{2F_2 - F_1 - F_1^2}.$$

$$\left[\text{ইঙ্গিত: } \bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i\,(x_i - \overline{x_k + w}) + x_k + w\right.$$

$$\left.\text{এবং } s^2 = \frac{1}{n}\sum_{i=1}^{k} f_i\,(x_i - \overline{x_k + w})^2 - \left\{\frac{1}{n}\sum_{i=1}^{k} f_i\,(x_i - \overline{x_k + w})\right\}^2\right]$$

**টীকা।** আলোচ্য ফল-দুটি ব্যবহার ক'রে সমদৈর্ঘ্য শ্রেণীবিন্যাসের ক্ষেত্রে অপেক্ষাকৃত অল্প আয়াসে ক্রমযৌগিক পরিসংখ্যার সাহায্যে গড় ও প্রমাণবিচ্যুতি নির্ণয় করা যায়। এই পদ্ধতিতে 3.7 সারণীতে প্রদত্ত রাশিতথ্যের গড় ও প্রমাণবিচ্যুতি নির্ণয় কর।

5.7 যদি $\bar{x}_i = \sum_{j=1}^{i} x_j/i$ হয়, [ $\bar{x}_1$, $\bar{x}_2$,........এগুলিকে **প্রগতি-গড়** (progressive mean) বলে ] তাহলে দেখাও যে

$$ns^2 = \sum_{i=2}^{n} \frac{i}{i-1}(x_i - \bar{x}_i)^2.$$

5.8 যথাক্রমে $f_1, f_2, \ldots, f_k$ পরিসংখ্যা সম্পন্ন $x_1, x_2, \ldots x_k$ এই $k$টি মানের গাণিতিক, গুণোত্তর ও প্রতিগাণিতিক গড় এবং প্রমাণবিচ্যুতি যথাক্রমে $A, G, H$ ও $s$ দ্বারা সূচিত হলে, যদি $k \geqslant 3$ এর জন্য প্রতিটি $\left(\frac{x_i - A}{A}\right)^k$-এর মান নগণ্য বিবেচনায় উপেক্ষণীয় হয়, তাহলে প্রমাণ কর :

(i) $G \simeq A\left(1 - \frac{s^2}{2A^2}\right)$ (iv) $AH \simeq G^2$

(ii) $H \simeq A(1 - s^2/A^2)$ (v) $A - 2G + H \simeq 0$

(iii) $A^2 - G^2 \simeq s^2$ (vi) $\frac{1}{n}\sum_{i=1}^{k} f_i \sqrt{x_i} \simeq \sqrt{A}\left(1 - \frac{s^2}{8A^2}\right)$

5.9 সহজেই প্রমাণ করা যায় $|A - x| + |B - x|$ লঘিষ্ঠ হবে যদি $A \leqslant x \leqslant B$ হয়। এই ফলটি ব্যবহার ক'রে (5.15)-এ প্রদত্ত অসমতা-সম্পর্কটির একটি বিকল্প প্রমাণ দাও।

5.10 সাংকেতিক চিহ্নগুলি স্বাভাবিক অর্থবাহী ধ'রে নিয়ে প্রমাণ কর,

$$MD_{\bar{x}} = \frac{2}{n}\left\{ \sum_{x_i < \bar{x}} f_i - \sum_{x_i < \bar{x}} f_i x_i \right\}$$

[ ইঙ্গিত : $\Sigma f_i (x_i - \bar{x}) = 0$ ]

এই ফলটি ব্যবহার ক'রে 3.7 সারণীতে প্রদত্ত রাশিতথ্যের গড়কেন্দ্রিক গড়-বিচ্যুতি নির্ণয় কর।

5.11 12 মাইল দীর্ঘ একটি রাস্তার ওপরে বিভিন্ন স্থানে একটি তেল কোম্পানীর কয়েকটি ন্যায্যমূল্যের কেরোসিন ডিপো আছে। কোম্পানী এখন ঐ রাস্তায় একটি কেন্দ্রীয় মজুতাগার স্থাপন ক'রে সেখান থেকে 1,000 লিটার তেল ধরে এমন একটি গাড়ীর সাহায্যে ডিপোগুলিতে তেল সরবরাহ করতে

চায়। নীচে ডিপোগুলির রাস্তার একপ্রান্ত থেকে দূরত্ব ও সাপ্তাহিক চাহিদা দেওয়া হ'ল :

| ডিপো | $A$ | $B$ | $C$ | $D$ | $E$ | $F$ | $G$ |
|---|---|---|---|---|---|---|---|
| রাস্তার একপ্রান্ত থেকে দূরত্ব ( মাইলে ) | $1\frac{1}{2}$ | 2 | 4 | 6 | $8\frac{1}{2}$ | 10 | 11 |
| সাপ্তাহিক চাহিদা ( হাজার লিটারে ) | 4 | 1 | 3 | 3 | 7 | 7 | 3 |

রাস্তার ঠিক কোন্ স্থানে প্রস্তাবিত মজুতাগারটি স্থাপন করা কোম্পানীর পক্ষে সুবিধাজনক হবে? সব সময় ভর্তি গাড়ীতে তেল পাঠানো হবে ধ'রে নাও।

5.12 3.7 ও 3.8 অনুশীলনীতে প্রদত্ত রাশিতথ্য থেকে পাওয়া পরিসংখ্যা-বিভাজনের জন্য প্রসার, গড়- ও মধ্যমাকেন্দ্রিক গড়বিচ্যুতি, প্রমাণবিচ্যুতি, চতুর্থক পার্থক্য, ভেদাঙ্ক ও চতুর্থক বিচ্যুতি-অঙ্ক নির্ণয় কর।

5.13 4.13 অনুশীলনীতে প্রদত্ত রাশিতথ্যের প্রমাণবিচ্যুতি নির্ণয় কর।

[ ইঙ্গিত : এখানে বিভাজনটিকে প্রথমে দুটি গোষ্ঠীতে ভাগ কর যেন প্রতি গোষ্ঠীর অন্তর্গত শ্রেণীগুলি সমদৈর্ঘ্য হয়। তারপর 5.9 সূত্র ব্যবহার কর। 4.13 অনুশীলনীতে প্রদত্ত ইঙ্গিত অনুসারেও প্রমাণবিচ্যুতি নির্ণয় করা যায়। ]

5.14 নীচের সারণীতে 8টি শহরের উপার্জনক্ষম ব্যক্তিদের সংখ্যা এবং মাসিক আয়ের গড় ও প্রমাণবিচ্যুতি দেওয়া হয়েছে। সামগ্রিকভাবে 8টি

| শহর | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| উপার্জনক্ষম ব্যক্তিদের সংখ্যা ( হাজারে ) | 28 | 56 | 128 | 146 | 220 | 216 | 140 | 56 |
| মাসিক আয়ের গড় ( টাকায় ) | 67·86 | 72·14 | 81·87 | 84·80 | 85·73 | 90·92 | 95·97 | 105·00 |
| মাসিক আয়ের প্রমাণ-বিচ্যুতি ( টাকায় ) | 10·12 | 13·81 | 15·95 | 16·64 | 16·01 | 17·22 | 14·78 | 8·86 |

শহরের উপার্জনক্ষম ব্যক্তিদের মাসিক আয়ের গড় ও প্রমাণবিচ্যুতি নির্ণয় কর।

5.15 নীচের সারণীতে একটি সমদৈর্ঘ্য শ্রেণীসম্পন্ন পরিসংখ্যা-বিভাজন দেওয়া হয়েছে। এখানে একটি যথেচ্ছ-গৃহীত মূলবিন্দু থেকে শ্রেণীদৈর্ঘ্যের এককে শ্রেণী-মধ্যকগুলির দূরত্ব $y$ দ্বারা সূচিত হচ্ছে। জানা আছে, বিভাজনটির গড় ও প্রমাণবিচ্যুতি যথাক্রমে 40·604 ও 7·92 একক। বিভাজনটি আয়তলেখ-এর সাহায্যে উপস্থাপিত কর।

| $y$ | −4 | −3 | −2 | −1 | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|---|---|---|
| $f$ | 3 | 15 | 45 | 57 | 50 | 36 | 25 | 9 |

5.16 কোন পরীক্ষায় 250 জন ছাত্রের প্রাপ্ত নম্বরের গড় ও প্রমাণবিচ্যুতি নির্ণয় করা হ'ল যথাক্রমে 45·50 ও 10·65. পরে ধরা পড়ল 5টি ছাত্রের নম্বর 41, 73, 28, 67 ও 33-এর পরিবর্তে ভুলক্রমে যথাক্রমে 47, 75, 20, 61 এবং 53 লেখা হয়েছিল। এই ভুল সংশোধন ক'রে গড় ও প্রমাণবিচ্যুতির সঠিক মান নির্ণয় কর।

[ ইঙ্গিত : মনে কর $x_1, x_2, \ldots, x_n$ এই $n$টি মানের মধ্যে $x_1, x_2, \ldots, x_k$ $(k < n)$টি মান অশুদ্ধ ব'লে ধরা পড়েছে। এদের শুদ্ধরূপ যথাক্রমে $x_1', \ldots, x_k'$ হলে, অশুদ্ধ গড় ও প্রমাণবিচ্যুতি $\bar{x}$ ও $s$ থেকে এদের শুদ্ধ রূপ $\bar{x}'$ ও $s'$ নিম্নলিখিতভাবে পাওয়া যায় :

$$n\bar{x}' = n\bar{x} - \sum^{k} x_i + \sum^{k} x_i' \text{ এবং } ns'^2 = \sum^{k} x_i'^2 + \sum_{k+1}^{n} x_i^2 - n\bar{x}'^2$$

$$= \{ n(s^2 + \bar{x}^2) - \sum_{1}^{n} x_i^2 + \sum_{1}^{n} x_i'^2 \} - n\bar{x}'^2. ]$$

5.17 1974 সালে কলকাতা সিনিয়র ডিভিশন ফুটবল লীগে ভ্রাতৃসঙ্ঘ, খিদিরপুর ও মহঃস্পোর্টিং ক্লাবের নীট গোলসংখ্যার ( একটি ম্যাচে নীট গোল-সংখ্যা = স্বপক্ষে গোল—বিপক্ষে গোল ) পরিসংখ্যা-বিভাজন নীচে দেওয়া হ'ল।

গোলসংখ্যার ভিত্তিতে কোন্ দলটি অধিকতর কৃতিত্বের দাবীদার, একটি উপযুক্ত মাপকের সাহায্যে বিচার কর।

| নীট গোলসংখ্যা | ম্যাচের সংখ্যা | | |
|---|---|---|---|
| | ভ্রাতৃসঙ্ঘ | খিদিরপুর | মহঃস্পোর্টিং |
| −2 | 2 | 1 | 0 |
| −1 | 4 | 2 | 4 |
| 0 | 6 | 11 | 10 |
| 1 | 2 | 2 | 3 |
| 2 | 4 | 2 | 2 |
| 3 | 0 | 1 | 0 |
| 4 | 1 | 0 | 0 |
| মোট | 19 | 19 | 19 |

## 5.12 নির্দেশিকা

1. Cook, L. H. L. *Statistical Problems and How to Solve Them.* Barnes & Noble, 1971.

2. Goon, A. M., Gupta M. K. & Dasgupta, B. *Fundamentals of Statistics.* Vol I. World Press, 1975.

3. Kenney, J. F., & Keeping E. S. *Mathematics of Statistics.* Part I. Van Nostrand, 1954.

4. Mounsey, J. *Introduction to Statistical Calculations.* English University Press, 1952.

5. Yule, G. U, & Kendall, M. G. *Introduction to the Theory of Statistics*, Charles Griffin, 1968.

# 6 পরিঘাত এবং প্রতিবৈষম্য- ও তীক্ষ্ণতা-মাপক
## ( Moments and Measures of Skewness and Kurtosis )

**6.1 পরিঘাতের সংজ্ঞা** (definition of moments) :

একাধিক পরিসংখ্যা-বিভাজনের মধ্যে তুলনা করার প্রয়োজন হলে আমরা এ পর্যন্ত বিভাজনগুলির মধ্যগামিতা এবং বিস্তৃতির বিচারে এই তুলনা করেছি। কিন্তু কেবলমাত্র এই দুটি বৈশিষ্ট্যের বিচারে তুলনা অনেক সময় পর্যাপ্ত হয় না, একাধিক পরিসংখ্যা-বিভাজনের গড় এবং প্রমাণবিচ্যুতি অভিন্ন হওয়া সত্ত্বেও বিভাজনগুলি প্রকৃতিতে ভিন্ন হতে পারে। এইসব ক্ষেত্রে তুলনার প্রয়োজনে বিভাজনগুলির আরও কিছু কিছু বৈশিষ্ট্যের সন্ধান করতে হয়—**প্রতিবৈষম্য** (skewness) এবং **তীক্ষ্ণতা** (kurtosis) হ'ল পরিসংখ্যা-বিভাজনের এইরকম দুটি বৈশিষ্ট্য। বৈশিষ্ট্য-দুটি এবং এদের মাপক সম্বন্ধে বিস্তারিত আলোচনা করার আগে আর এক ধরনের বিবরণাত্মক মাপকের সঙ্গে আমাদের পরিচিতি হতে হবে —এটি হ'ল **পরিঘাত** (moment)।

কোনও চলের একাধিক মান প্রদত্ত হলে একটি বিশেষ বিন্দু $A$ থেকে মানগুলির বিচ্যুতির $r$-তম ঘাতের গড়কে বলা হয় চলটির ( বা পরিসংখ্যা বিভাজনটির ) **r-তম A-কেন্দ্রিক পরিঘাত** ($r$th moment about $A$)। এটি সাধারণতঃ $m'_r$ দ্বারা নির্দেশ করা হয়, অর্থাৎ,

$$m'_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - A)^r. \qquad \cdots \quad (6.1)$$

$A=\bar{x}$ হলে পাওয়া যায় **r-তম গড়কেন্দ্রিক পরিঘাত** ($r$th central moment) $m_r$, অর্থাৎ,

$$m_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r. \qquad \cdots \qquad \cdots \quad (6.2)$$

শ্রেণীবিন্যস্ত রাশিতথ্যের ক্ষেত্রে

$$m'_r = \frac{1}{n}\sum_{i=1}^{n} f_i(x_i - A)^r, \qquad \cdots \qquad \cdots \quad (6.1a)$$

এবং $$m_r = \frac{1}{n}\sum_{i=1} f_i(x_i - \bar{x})^r. \qquad (6.2a)$$

(6.1) ও (6.1$a$) সূত্রে $A=0$ নেওয়া হলে পাওয়া যায় **শূন্যকেন্দ্রিক পরিঘাত** (moment about zero)। $m'_r$-কে অনেক সময় **অশোধিত পরিঘাত** (raw moment)ও বলা হয়।

ওপরের সূত্রগুলিতে $r$-এর মান শূন্য বা যে কোন অখণ্ড ধনসংখ্যা হতে পারে। অশোধিত এবং গড়কেন্দ্রিক পরিঘাতের প্রতীকচিহ্ন-দুটির পার্থক্য লক্ষণীয়।

আর এক ধরনের পরিঘাত হ'ল **গৌণিক পরিঘাত** (factorial moment)। $r$-তম গৌণিক পরিঘাত $m'_{[r]}$-এর সংজ্ঞা হ'ল

$$m'_{[r]} = \begin{cases} \frac{1}{n}\sum_{i=1}^{n} x_i(x_i-1)\cdots(x_i-r+1) \\ \frac{1}{n}\sum_{i=1}^{k} f_i x_i(x_i-1)\cdots(x_i-r+1) \end{cases} \qquad \cdots \quad (6.3)$$

পরিসংখ্যা-বিভাজনের ক্ষেত্রে সাধারণতঃ গৌণিক পরিঘাত নির্ণয় করা হয় না। তবে বিচ্ছিন্ন চল সংক্রান্ত তত্ত্বগত নিবেশনের ( অষ্টম পরিচ্ছেদ দ্রষ্টব্য ) ক্ষেত্রে এগুলির বহুল ব্যবহার রয়েছে।

**r-তম চিহ্ননিরপেক্ষ পরিঘাতের** ($r$th absolute moment) সংজ্ঞা হ'ল :

$$n'_r = \begin{cases} \frac{1}{n}\sum_{i=1}^{n} |x_i - A|^r \\ \frac{1}{n}\sum_{i=1}^{k} f_i |x_i - A|^r \end{cases} \qquad \cdots \qquad \cdots \quad (6.4)$$

এবং $$n_r = \begin{cases} \frac{1}{n}\sum_{i=1}^{n} |x_i - \bar{x}|^r \\ \frac{1}{n}\sum_{i=1}^{k} f_i |x_i - \bar{x}|^r \end{cases} \qquad \cdots \qquad \cdots \quad (6.5)$$

লক্ষ্য কর, $m'_1$ এবং $m_2$ যথাক্রমে গাণিতিক গড় এবং ভেদমান অর্থাৎ প্রমাণবিচ্যুতির বর্গ, যেগুলির সঙ্গে আমরা ইতিপূর্বেই পরিচিত হয়েছি।

স্পষ্টতঃই, যে কোন চলের ক্ষেত্রেই

$$m'_0 = m_0 = 1 \text{ এবং } m_1 = 0 \quad \cdots \quad \cdots \quad (6.6)$$

## 6.2 রৈখিক রূপান্তর এবং গড়কেন্দ্রিক পরিঘাত :

মনে কর, $y_i = \frac{x_i - a}{c}$, $i = 1\,(1)\,k$.

সুতরাং $y$-এর $r$-তম গড়কেন্দ্রিক পরিঘাত

$$m_r(y) = \frac{1}{n}\sum_{i=1}^{k} f_i\,(y_i - \bar{y})^r = \frac{1}{n}\sum_{i=1}^{k} f_i\left(\frac{x_i - a}{c} - \frac{\bar{x} - a}{c}\right)^r,$$

$$= \frac{1}{c^r}\cdot\frac{1}{n}\sum_{i=1}^{k} f_i\,(x_i - \bar{x})^r = \frac{1}{c^r}\cdot m_r(x). \quad \cdots \quad (6.7)$$

সুতরাং দেখা যাচ্ছে, গড়কেন্দ্রিক পরিঘাতের ওপর মূলবিন্দু পরিবর্তনের কোনও প্রভাব নেই। এই ধর্মটি গড়কেন্দ্রিক পরিঘাত নির্ণয়নে খুবই সহায়ক।

## 6.3 গড়কেন্দ্রিক পরিঘাত এবং অশোধিত পরিঘাতের মধ্যে সম্পর্ক :

$r$-এর যে কোন মানের জন্য $r$-তম গড়কেন্দ্রিক পরিঘাতকে প্রথম, দ্বিতীয়, $\cdots$, $r$-তম $A$-কেন্দ্রিক পরিঘাতের সাহায্যে প্রকাশ করা যায়। মনে কর,

$$m'_r = \frac{1}{n}\sum_{i=1}^{k} f_i\,(x_i - A)^r.$$

সুতরাং $m'_1 = \frac{1}{n}\sum_{i=1}^{k} f_i\,(x_i - A) = \bar{x} - A. \quad \cdots \quad (6.8)$

এখন, $m_r = \frac{1}{n}\sum_{i=1}^{k} f_i\,(x_i - \bar{x})^r$

$$\frac{1}{n}\sum_{i=1}^{n} f_i\{(x_i-A)-(\bar{x}-A)\}^r$$

$$=\frac{1}{n}\sum_i f_i\{(x_i-A)-m'_1\}^r$$

$$=\frac{1}{n}\sum_i f_i\{(x_i-A)^r-\binom{r}{1}(x_i-A)^{r-1}m'_1$$
$$+\binom{r}{2}(x_i-A)^{r-2}m'^2_1-\cdots+(-1)^r m'^r_1\}$$

$$=m'_r-\binom{r}{1}m'_{r-1}m'_1+\binom{r}{2}m'_{r-2}m'^2_1-\cdots+(-1)^r m'^r_1. \quad (6.9)$$

সাধারণতঃ, $m_2$, $m_3$ এবং $m_4$ এই তিনটি গড়কেন্দ্রিক পরিঘাতই বেশী ব্যবহৃত হয়। (6.9) সূত্রে $r=2, 3, 4$ বসিয়ে পাওয়া যায়,

$$\begin{aligned}
m_2&=m'_2-2m'_1m'_1+m'^2_1\\
&=m'_2-m'^2_1,\\
m_3&=m'_3-3m'_2m'_1+3m'_1m'^2_1-m'^3_1\\
&=m'_3-3m'_2m'+2m'^3_1,\\
m_4&=m'_4-4m'_3m'_1+6m'_2m'^2_1-4m'_1m'^3_1+m'^4_1\\
&=m'_4-4m'_3m'_1+6m'_2m'^2_1-3m'^4_1.
\end{aligned} \quad (6.10)$$

অনুরূপভাবে $r$-তম $A$-কেন্দ্রিক পরিঘাতও $r$ এবং ক্ষুদ্রতর ক্রমের গড়কেন্দ্রিক পরিঘাতের সাহায্যে প্রকাশ করা যায়।

$$m'_r=\frac{1}{n}\sum_{i=1}^{k} f_i(x_i-A)^r$$

$$=\frac{1}{n}\sum_i f_i\{(x_i-\bar{x})+(\bar{x}-A)^r\}$$

$$=\frac{1}{n}\sum_i f_i\{(x_i-\bar{x})^r+\binom{r}{1}(x_i-\bar{x})^{r-1}d+\binom{r}{2}(x_i-\bar{x})^{r-2}d^2$$
$$+\cdots+d^r\}, \text{ যেখানে } d=\bar{x}-A$$

$$=m_r+\binom{r}{1}m_{r-1}d+\binom{r}{2}m_{r-2}d^2+\cdots+d^r. \quad \cdots \quad (6.11)$$

(6.11) সূত্রে $r=2, 3, 4$ বসিয়ে পাওয়া যায়

$$\left.\begin{aligned} m'_2 &= m_2 + d^2 \\ m'_3 &= m_3 + 3m_2 d + d^3 \\ m'_4 &= m_4 + 4m_3 d + 6m_2 d^2 + d^4, \end{aligned}\right\} \qquad (6.12)$$

যেহেতু $m_1 = 0,\ m_0 = 1.$

## 6.4 পরিঘাত নির্ণয়ন-পদ্ধতি :

সাধারণতঃ কোন পরিসংখ্যা-বিভাজনের $m'_1$, $m_2$, $m_3$ এবং $m_4$—এই চারটি পরিঘাতেরই বেশী ব্যবহার দেখা যায়। এগুলির মধ্যে $m'_1$ এবং $m_2$-এর নির্ণয়ণ পদ্ধতি আগেই আলোচিত হয়েছে।

সাধারণভাবে গড়কেন্দ্রিক পরিঘাত নির্ণয়ের উদ্দেশ্যে প্রথমে সুবিধামত কোন মূলবিন্দুকে কেন্দ্র ক'রে পরিঘাত নির্ণয় করা হয় এবং পরে (6.10) সূত্রটি ব্যবহার ক'রে পাওয়া যায় প্রয়োজনীয় গড়কেন্দ্রিক পরিঘাত [ উদা. 6.1 ]।

গোষ্ঠীবদ্ধ রাশিতথ্যের ক্ষেত্রে শ্রেণীগুলি সমদৈর্ঘ্য হলে মাপনামাত্রার পরিবর্তন সাধন ক'রে (6.7) সূত্র ব্যবহার করলেও পরিঘাত-নির্ণয়নে শ্রম আরও কিছুটা লাঘব হতে পারে [ উদা. 6.1 ]।

সর্বাধিক শ্রমসঙ্কোচের উদ্দেশ্যে অশোধিত পরিঘাতের মূলকেন্দ্র হিসাবে নির্বাচন করা হয় গৃহীত প্রসারের মাঝামাঝি কোন মানকে। সমদৈর্ঘ্য শ্রেণী-বিন্যাসের ক্ষেত্রে সাধারণতঃ মধ্যবর্তী শ্রেণীর মধ্যকটিকে মূলকেন্দ্র এবং সাধারণ শ্রেণী-দৈর্ঘ্যটিকে মাপনামাত্রা হিসাবে নেওয়া হয়ে থাকে।

পরিঘাত নির্ণয়নে শুদ্ধিবিচারের জন্য **শার্লিয়ারের শুদ্ধি পরীক্ষার** (Charlier's check) সূত্র ব্যবহার করা যেতে পারে। লক্ষ্য কর :

$$\sum_{i=1}^{k} f_i(x_i+1)^r = \sum_{i=1}^{k} f_i x_i^r + \binom{r}{1}\sum_{i=1}^{k} f_i x_i^{r-1} + \cdots$$

$$\cdot \binom{r}{r-1}\sum_{i=1}^{k} f_i x_i + n. \quad \cdots \qquad (6.13)$$

(6.13) সূত্রটিই শার্লিয়ারের শুদ্ধি পরীক্ষার সূত্র। যে কোন ক্রমের

পরিঘাতের নির্ণীত মানের শুদ্ধি পরীক্ষার জন্য $r$-এর সংশ্লিষ্ট মানটি (6.13) সূত্রে বসানো চলে। যেমন, $r=2, 3, 4$ হলে, যথাক্রমে

$$\left.\begin{aligned}
\sum_i f_i(x_i+1)^2 &= \sum_i f_i x_i^2 + 2\sum_i f_i x_i + n,\\
\sum_i f_i(x_i+1)^3 &= \sum_i f_i x_i^3 + 3\sum_i f_i x_i^2 + 3\sum_i f_i x_i + n,\\
\sum_i f_i(x_i+1)^4 &= \sum_i f_i x_i^4 + 4\sum_i f_i x_i^3 + 6\sum_i f_i x_i^2 + 4\sum_i f_i x_i + n.
\end{aligned}\right\} \cdots \quad (6.14)$$

এখন কোন উচ্চতর ক্রমের পরিঘাত নির্ণয়ের সঙ্গে সঙ্গে সাধারণতঃ অধঃক্রমিক পরিঘাতগুলিও নির্ণয় করা হয়। সুতরাং পরিঘাত নির্ণয়ের জন্য ব্যবহৃত ছকে অতিরিক্ত আর একটি স্তম্ভ ব্যবহার ক'রে সহজেই সূত্রটির সাহায্যে নির্ণীত মানগুলির শুদ্ধিবিচার করা যেতে পারে।

**উদা. 6.1** গত বছরের সেকেণ্ডারী পরীক্ষায় 1,000 জন ছাত্রের অঙ্কের শতকরা নম্বরের পরিসংখ্যা-বিভাজন নীচে দেওয়া হ'ল ( কাল্পনিক রাশিতথ্য )।

**সারণী 6.1**

**1,000 জন ছাত্রের অঙ্কের শতকরা নম্বরের পরিসংখ্যা-বিভাজন**

| নম্বর (%) | ছাত্রসংখ্যা |
|---|---|
| 1—10 | 2 |
| 11—20 | 6 |
| 21—30 | 29 |
| 31—40 | 108 |
| 41—50 | 447 |
| 51—60 | 240 |
| 61—70 | 121 |
| 71—80 | 42 |
| 81—90 | 4 |
| 91—100 | 1 |
| মোট | 1,000 |

এখানে বিভিন্ন ক্রমের পরিঘাত নির্ণয় করার জন্য নিম্নলিখিত ছকে অঙ্কপাতন করা হ'ল।

## সারণী 6.2

## 6.1 সারণীতে প্রদত্ত পরিসংখ্যা-বিভাজনের পরিঘাত নির্ণয়

| শ্রেণী-মধ্যক $x$ | $f$ | $y=\frac{x-45\cdot5}{10}$ | $yf$ | $y^2f$ | $y^3f$ | $y^4f$ | $(y+1)^4f$ |
|---|---|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| 5·5 | 2 | −4 | −8 | 32 | −128 | 512 | 162 |
| 15·5 | 6 | −3 | −18 | 54 | −168 | 486 | 96 |
| 25·5 | 29 | −2 | −58 | 116 | −232 | 464 | 29 |
| 35·5 | 108 | −1 | −108 | 108 | −108 | 108 | 0 |
| 45·5 | 447 | 0 | 0 | 0 | 0 | 0 | 447 |
| 55·5 | 240 | 1 | 240 | 240 | 240 | 240 | 3,840 |
| 65·5 | 121 | 2 | 242 | 484 | 968 | 1,936 | 9,801 |
| 75·5 | 42 | 3 | 126 | 378 | 1,134 | 3,402 | 10,752 |
| 85·5 | 4 | 4 | 16 | 64 | 256 | 1,024 | 2,500 |
| 95·5 | 1 | 5 | 5 | 25 | 125 | 625 | 1,296 |
| মোট | 1,000 | — | 437 | 1,501 | 2,093 | 8,797 | 28,927 |

**শার্লিয়ারের শুদ্ধি পরীক্ষা ঃ**

$\sum_i (y_i+1)^4 f_i = 28927.$ আবার, $\sum_i y_i^4 f_i +$

$$4\sum_i y_i^3 f_i + 6\sum_i y_i^2 f_i + 4\sum_i y_i f_i + n$$

$$= 8797 + 4\times2093 + 6\times1501 + 4\times437 + 1000$$

$$= 28927.$$

সুতরাং আমাদের অঙ্কপাতন ভ্রমশূন্য হয়েছে ব'লে ধ'রে নেওয়া যেতে পারে এখন, $y$-এর জন্য,

শূন্যকেন্দ্রিক পরিঘাত :

$$m'_1(y) = 437/1{,}000 = 0{\cdot}437$$
$$m'_2(y) = 1501/1{,}000 = 1{\cdot}501$$
$$m'_3(y) = 2093/1{,}000 = 2{\cdot}093$$
$$m'_4(y) = 8797/1{,}000 = 8{\cdot}797$$

সুতরাং $y$-এর গড়কেন্দ্রিক পরিঘাত :

$$m_2(y) = 1{\cdot}501 - 0{\cdot}437^2 = 1{\cdot}310031$$
$$m_3(y) = 2{\cdot}0930 - 3 \times 1{\cdot}501 \times 0{\cdot}437 + 2 \times (0{\cdot}437)^3$$
$$= 0{\cdot}292095$$

এবং
$$m_4(y) = 8{\cdot}797 - 4 \times 2{\cdot}093 \times 0{\cdot}437 + 6 \times 1{\cdot}501({\cdot}437)^2$$
$$- 3 \times (0{\cdot}437)^4 = 6{\cdot}748893$$

সুতরাং $x$-এর জন্য পরিঘাতগুলি হবে

$$m'_1(x) = 45{\cdot}5 + 10 . m'_1(y) = 49{\cdot}87$$
$$m_2(x) = 10^2 . m_2(y) = 131{\cdot}0031$$
$$m_3(x) = 10^3 . m_3(y) = 292{\cdot}0950$$
$$m_4(x) = 10^4 . m_4(y) = 67488{\cdot}9300.$$

## 6.5 শেপার্ডের (W. S. Sheppard) পরিঘাত সম্পর্কিত শুদ্ধি (corrections for grouping) :

পূর্বেই বলা হয়েছে, শ্রেণীবিন্যস্ত রাশিতথ্যের ক্ষেত্রে গাণিতিক গড় কিংবা প্রমাণবিচ্যুতি নির্ণয় করা হয় বিভিন্ন শ্রেণীর অন্তর্গত মানগুলি সংশ্লিষ্ট শ্রেণী-মধ্যকের সমান এই স্বীকরণসাপেক্ষে,—অর্থাৎ, শ্রেণী-পরিসংখ্যাগুলি শ্রেণী-মধ্যকের পরিসংখ্যা ধ'রে নিয়ে অবিচ্ছিন্ন চলের পরিসংখ্যা-বিভাজনটিকে প্রকৃতপক্ষে একটি বিচ্ছিন্ন চলের পরিসংখ্যা-বিভাজনে পর্যবসিত করা হয়। সাধারণভাবে যে কোন ক্রমের পরিঘাত নির্ণয়নের ক্ষেত্রেই এই নীতি অনুসরণ করা হয়ে থাকে। স্পষ্টতঃই এই স্বীকরণসাপেক্ষে পরিঘাতের নির্ণীত মানে কিছুটা ভ্রান্তি থাকা স্বাভাবিক। এই ভ্রান্তিকে বলা হয় **গোষ্ঠীবন্ধন ভ্রান্তি** (error due to grouping)। **শেপার্ড** প্রদত্ত শুদ্ধি প্রয়োগ ক'রে পরিঘাতের নির্ণীত মান থেকে এই গোষ্ঠীবন্ধন ভ্রান্তি অনেকখানি দূর করা যায়।

$m'_r$ দ্বারা সংশোধিত এবং $\overline{m}'_r$ দ্বারা অসংশোধিত পরিঘাত সূচিত ক'রে প্রথম চারটি যথেচ্ছগৃহীতমূল-কেন্দ্রিক পরিঘাতের জন্য শুদ্ধিগুলি হ'ল :

$$\left.\begin{aligned} m'_1 &= \overline{m}'_1 \\ m'_2 &= \overline{m}'_2 - \frac{w^2}{12} \\ m'_3 &= \overline{m}'_3 - \frac{w^2}{4}\overline{m}'_1 \\ m'_4 &= \overline{m}'_4 - \frac{w^2}{2}\overline{m}'_2 + \frac{7}{240}w^4, \end{aligned}\right\} \quad \cdots \quad (6.15)$$

$w$ = শ্রেণীদৈর্ঘ্য ( প্রতিটি শ্রেণীর জন্য সমান ধ'রে )।

তেমনি অনুরূপ প্রতীক ব্যবহারে গড়কেন্দ্রিক পরিঘাতের জন্য শুদ্ধি :

$$\left.\begin{aligned} m_2 &= \overline{m}_2 - \frac{w^2}{12} \\ m_3 &= \overline{m}_3 \\ m_4 &= \overline{m}_4 - \frac{w^2}{2}\overline{m}_2 + \frac{7}{240}w^4 \end{aligned}\right\} \quad \cdots \quad (6.16)$$

অনুরূপভাবে গৌণিক পরিঘাতের জন্য ওয়াল্ড (Wald) নিম্নলিখিত শুদ্ধির ব্যবস্থা দিয়েছেন :

$$\left.\begin{aligned} m'_{[1]} &= m'_{[1]} \\ m'_{[2]} &= m'_{[2]} - \frac{w^2}{12} \\ m'_{[3]} &= m'_{[3]} - \frac{w^2}{4}m'_{[1]} + \frac{w^3}{4} \\ m'_{[4]} &= m'_{[4]} - \frac{w^2}{2}m'_{[2]} + w^3 m'_{[1]} - \frac{71}{80}w^4. \end{aligned}\right\} \quad \cdots \quad (6.17)$$

যে কোন পরিসংখ্যা-বিভাজন থেকে নির্ণীত পরিঘাতের ক্ষেত্রে কিন্তু শুদ্ধিগুলি প্রয়োগ করা যুক্তিসঙ্গত হবে না। যে সব স্বীকরণের ভিত্তিতে শেপার্ড আলোচ্য শুদ্ধিগুলি পেয়েছিলেন সেগুলি হ'ল প্রথমতঃ, সংশ্লিষ্ট পরিসংখ্যা-বিভাজনটি হবে কোন অবিচ্ছিন্ন চলের। দ্বিতীয়তঃ, চলের পরিসংখ্যা রেখাটিকে X-অক্ষের সঙ্গে **উচ্চক্রমিক সংযোগ** (high order contact)-সম্পন্ন হতে হবে প্রসারের উভয় দিকেই—অর্থাৎ পরিসংখ্যা-রেখাটিকে শুরু এবং শেষ উভয় দিকেই X-অক্ষের সঙ্গে ক্রমাসন্ন (asymptotic) হতে হবে। সুতরাং প্রদত্ত কোন পরিসংখ্যা-বিভাজন

সম্পর্কে ওপরের স্বীকরণ-দুটি সত্য হলে তবেই সেক্ষেত্রে শুদ্ধিগুলি প্রয়োগ করা যাবে। সসীমসংখ্যক রাশিতথ্য থেকে দ্বিতীয় সর্তটি যাচাই করা স্পষ্টতঃই খুব কঠিন—এক্ষেত্রে মোট পরিসংখ্যা মোটামুটি বেশী হলে পরিসংখ্যা-বিভাজনে শ্রেণী-পরিসংখ্যাগুলি প্রসারের উভয়প্রান্তে ক্রমশঃ কমতে কমতে শূন্যের কাছাকাছি পৌঁছেছে দেখা গেলে সর্তটি মোটামুটিভাবে পালিত হয়েছে ধ'রে নেওয়া হয়।

উল্লিখিত সর্ত-দুটি ছাড়াও আরও দুটি সর্ত পালিত হওয়া উচিত, নয়তো শুদ্ধিগুলির প্রয়োগ অর্থহীন হয়ে পড়বে। প্রথমতঃ, সাধারণ শ্রেণীদৈর্ঘ্য খুব কম (অর্থাৎ শ্রেণীসংখ্যা খুব বেশী) হলে শুদ্ধিগুলি প্রয়োগ করা উচিত নয়—কারণ সেক্ষেত্রে শুদ্ধিগুলির ফল খুবই নগণ্য দাঁড়াবে। দ্বিতীয়তঃ, মোট পরিসংখ্যা খুব কম হলেও শুদ্ধিগুলি প্রয়োগ করা উচিত নয়—কারণ সেক্ষেত্রে পরিসংখ্যা-বিভাজন থেকে পাওয়া পরিঘাতগুলিতে গোষ্ঠীবদ্ধন ভ্রান্তি অপেক্ষা নমুনাজ ভ্রান্তির পরিমাণ বেশী হয়ে দাঁড়াবে। এ সম্বন্ধে সাধারণ নিয়ম হ'ল, কোন পরিসংখ্যা-বিভাজনের মোট পরিসংখ্যা 1,000 বা তার বেশী হলে এবং মোট শ্রেণীসংখ্যা 20 অথবা তার কম হলে তবেই বিভাজনটির ক্ষেত্রে শুদ্ধিগুলির প্রয়োগ অনুমোদন করা যায়। অবশ্য যেসব ক্ষেত্রে শুদ্ধিগুলি প্রযোজ্য সেখানে পরিঘাত নির্ণয়ে এদের প্রয়োগ আবশ্যিক।

**উদা. 6.2** উদা 6.1-এ প্রদত্ত পরিসংখ্যা-বিভাজনের ক্ষেত্রে ওপরের সবকটি সর্তই পালিত হয়েছে। সুতরাং এক্ষেত্রে শুদ্ধিগুলি প্রয়োগ করা যাক। শোধিত পরিঘাতগুলি দাঁড়াবে

$$m_2 = 131{\cdot}0031 - 100/12 = 122{\cdot}6698$$

$$m_4 = 67488{\cdot}9300 - 131{\cdot}0031 \times \frac{10^2}{2} + 10^4 \times \frac{7}{240}$$

$$= 60967{\cdot}9400.$$

## 6.6 প্রতিবৈষম্য এবং প্রতিবৈষম্য-মাপক (skewness and its measures):

আগেই বলা হয়েছে **প্রতিবৈষম্য** পরিসংখ্যা-বিভাজনের তথা চলের একটি গুরুত্বপূর্ণ বৈশিষ্ট্য। 4.6 অনুচ্ছেদে আমরা এই বৈশিষ্ট্যটির স্বরূপ কিছুটা আলোচনা করেছি। কোনও বিভাজনের প্রতিসম অবস্থা থেকে বিচ্যুতির মাত্রাই হল বিভাজনটির প্রতিবৈষম্য—এই মাত্রাটি পরিমাপ করার জন্য বিভিন্ন মাপকের কথা ভাবা যেতে পারে।

4.6 অনুচ্ছেদে আলোচিত প্রতিসম এবং দক্ষিণায়ত ও বামায়ত প্রতিবিষম

বিভাজনের ক্ষেত্রে গাণিতিক গড়, মধ্যমা এবং ভূয়িষ্ঠকের আপেক্ষিক অবস্থিতি থেকে দেখা যায় $\bar{x}-\breve{x}$-এর মান যথাক্রমে শূন্য, ধনাত্মক এবং ঋণাত্মক। স্বভাবতঃই,

$$SK_1 = \frac{\bar{x}-\breve{x}}{s} \qquad \cdots \quad (6.18)$$

প্রকাশনটিকে প্রতিবৈষম্য-মাপক হিসাবে ব্যবহার করা যেতে পারে। $\bar{x}-\breve{x}$ কে $s$-এর অনুপাতে প্রকাশ করার উদ্দেশ্য হ'ল মাপকটিকে একক-নিরপেক্ষ একটি শুদ্ধ সংখ্যায় প্রকাশ করা। $\breve{x}$-এর মান সহজে পাওয়া না গেলে (4.13) সূত্রে প্রদত্ত অবেক্ষণভিত্তিক সম্পর্কটি ব্যবহার ক'রে

$$SK_2 = \frac{3(\bar{x}-\tilde{x})}{s} \qquad \cdots \quad (6.19)$$

এই দ্বিতীয় মাপকটি নেওয়া হয়। স্পষ্টতঃই (6.18) ও (6.19) সূত্রে $s > 0$ ধরা হয়েছে।

এখন $\left|\bar{x}-\tilde{x}\right| = \frac{1}{n}\left|\sum (x_i-\tilde{x})\right| \leqslant \frac{1}{n}\sum\left|x_i-\tilde{x}\right| \leqslant s$ ( 5.17 ফল অনুসারে ), সুতরাং $-3 \leqslant SK_2 \leqslant 3$. $SK_1$-এর মানসীমাও মোটামুটি $\pm 3$.

আর একটি প্রতিবৈষম্য-মাপক পাওয়া যায় প্রথম, দ্বিতীয় এবং তৃতীয় চতুর্থকের আপেক্ষিক অবস্থিতি থেকে। স্পষ্টতঃই প্রতিসম বিভাজনে $Q_1$ এবং $Q_3$ মধ্যমা $Q_2$ থেকে সমদূরবর্তী, দক্ষিণায়ত এবং বামায়ত প্রতিবিষম বিভাজনে যথাক্রমে $Q_1$ এবং $Q_3$ মধ্যমা $Q_2$-এর অধিকতর নিকটবর্তী। সুতরাং

$$SK_3 = \frac{(Q_3-Q_2)-(Q_2-Q_1)}{(Q_3-Q_2)+(Q_2-Q_1)} = \frac{Q_3-2Q_2+Q_1}{2Q} \quad (6.20)$$

এই প্রকাশনটি ( এখানে $Q$ = চতুর্থক বিচ্যুতি ) প্রতিবৈষম্য-মাপক হিসাবে ব্যবহৃত হতে পারে। স্পষ্টতঃই এটির মান একক-নিরপেক্ষ হবে। চরম দক্ষিণায়ত প্রতিবিষম বিভাজনের ক্ষেত্রে $Q_1 \simeq Q_2$ এবং চরম বামায়ত প্রতিবিষম বিভাজনের ক্ষেত্রে $Q_3 \simeq Q_2$. সুতরাং $SK_3$-এর মানসীমা $\pm 1$.

গড়কেন্দ্রিক পরিঘাতের একটি বৈশিষ্ট্য হ'ল, প্রতিসম এবং দক্ষিণায়ত ও বামায়ত প্রতিবিষম বিভাজনের ক্ষেত্রে অযুগ্মক্রমিক যে কোন গড়কেন্দ্রিক পরিঘাতের মান যথাক্রমে শূন্য, ধনাত্মক এবং ঋণাত্মক—অবশ্য $m_1$ একটি ব্যতিক্রম, কেননা যে কোন বিভাজনের ক্ষেত্রেই এটির মান সংজ্ঞানুসারেই শূন্য। সুতরাং $m_1$ ব্যতীত যে কোন অযুগ্মক্রমিক গড়কেন্দ্রিক পরিঘাতই প্রতিবৈষম্য-মাপক হিসাবে ব্যবহার করা যেতে পারে। নির্ণয়নে শ্রমলাঘব এবং নমুনাজ

বিচ্যুতির পরিমাণের বিচারে ( পরিঘাতের ক্রম বাড়ানোর সঙ্গে সঙ্গে নমুনাজ বিচ্যুতির পরিমাণও বাড়ে ; নমুনাজ বিচ্যুতি যথাসম্ভব কম হওয়াই বাঞ্ছনীয় ) স্বভাবতঃই পরবর্তী উচ্চতরক্রমিক অযুগ্ম পরিঘাত $m_3$-কে বেছে নেওয়া হয় এই উদ্দেশ্যে এবং এটিকে একক-নিরপেক্ষ ক'রে পাওয়া যায় আর একটি মাপক

$$g_1 = \frac{m_3}{s^3} \ (s > 0 \text{ স্বীকরণসাপেক্ষে })। \quad \cdots \quad (6.21)$$

অনেক সময় শুধুমাত্র প্রতিবৈষম্যের পরমমাত্রা নিরূপণের প্রয়োজনে $b_1 = g_1{}^2$ প্রকাশনটিও মাপক হিসাবে ব্যবহৃত হয়। $g_1$ ও $b_1$-এর মানসীমা যথাক্রমে $(-\infty, \infty)$ এবং $(0, \infty)$ হলেও বাস্তবক্ষেত্রে সাধারণতঃ এগুলির মান খুব বেশী হয় না।

## 6.7 তীক্ষ্ণতা এবং তীক্ষ্ণতা-মাপক (kurtosis and its measures) :

পরিসংখ্যা-বিভাজনের মধ্যগামিতা, বিস্তৃতি এবং প্রতিবৈষম্য সম্বন্ধে ধারণা পাওয়া গেলে সংশ্লিষ্ট চলটির বিভাজনের আকৃতি সম্বন্ধে কিছুটা চিত্র পাওয়া যায়, কিন্তু চিত্রটি সম্পূর্ণ পেতে হলে বিভাজনের আর একটি বৈশিষ্ট্য সম্বন্ধে জানতে হবে। এটি হ'ল বিভাজনের **তীক্ষ্ণতা**। চলের প্রদত্ত মানগুলির ভূয়িষ্ঠকের কাছাকাছি কেন্দ্রীভবনের মাত্রাকে বিভাজনটির তীক্ষ্ণতা আখ্যা দেওয়া যেতে পারে—এই মাত্রা যত বেশী, সংশ্লিষ্ট চলের পরিসংখ্যা-রেখাটির ( বিচ্ছিন্ন চলের ক্ষেত্রে কল্পিত ) শীর্ষদেশ ততই তীক্ষ্ণ। অভিন্ন গড়, প্রমাণবিচ্যুতি এবং প্রতিবৈষম্য-সম্পন্ন একাধিক বিভাজনের তীক্ষ্ণতার মাত্রা ভিন্ন হতে পারে।

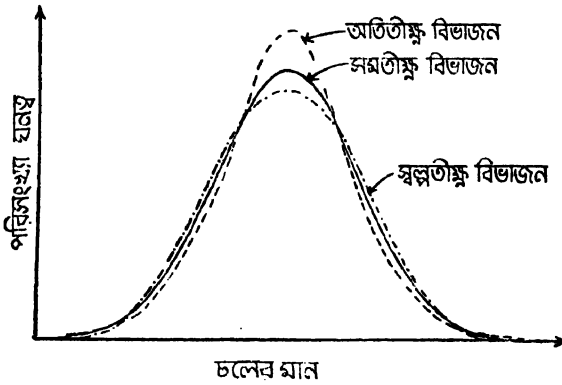
যে কোন যুগ্মক্রমিক গড়কেন্দ্রিক পরিঘাত বিভাজনের তীক্ষ্ণতা পরিমাপের জন্য ব্যবহার করা যেতে পারে। মাপকটিকে একক-নিরপেক্ষ করার জন্য সাধারণতঃ প্রমাণবিচ্যুতির ($s = \sqrt{m_2}$) এককে দেওয়া হয় ব'লে পরবর্তী উচ্চতরক্রমিক গড়-কেন্দ্রিক পরিঘাত $m_4$ সহযোগে তীক্ষ্ণতা-মাপক নেওয়া হয়।

$$g_2 = \frac{m_4}{s^4} - 3 \quad \cdots \quad (6.22)$$

একটি প্রচলিত তীক্ষ্ণতা-মাপক।

নর্ম্যাল বিভাজনের ( অষ্টম পরিচ্ছেদে আলোচিত হবে ) ক্ষেত্রে $m_4 = 3s^4$ সুতরাং $g_2 = 0$. প্রদত্ত কোন পরিসংখ্যা-বিভাজনের ক্ষেত্রে $g_2$-র মান শূন্য হওয়ার অর্থ আলোচ্য চলটির বিভিন্ন মানগুলির সংখ্যাগরিষ্ঠমানের কাছাকাছি

কেন্দ্রীভবনের মাত্রা, একই প্রমাণবিচ্যুতি-সম্পন্ন একটি নর্ম্যাল বিভাজনের এই মাত্রার সমান। এক্ষেত্রে তীক্ষ্ণতার মাত্রা স্বাভাবিক অপেক্ষা খুব বেশীও নয়, খুব কমও নয়—তাই সংশ্লিষ্ট বিভাজনটিকে **মধ্যমতীক্ষ্ণ** (mesokurtic) বলা হয়। পক্ষান্তরে $g_2$-র মান ধনাত্মক ও ঋণাত্মক হওয়ার অর্থ, আলোচ্য মাত্রা একটি সদৃশ নর্ম্যাল বিভাজনের তুলনায় যথাক্রমে বেশী ও কম। সুতরাং ধনাত্মক ও ঋণাত্মক $g_2$ সম্পন্ন বিভাজনকে যথাক্রমে বলা হয় **অতিতীক্ষ্ণ** (leptokurtic) এবং **স্বল্পতীক্ষ্ণ** (platykurtic)।

6.1 চিত্রে অভিন্ন গড় ও প্রমাণবিচ্যুতি-সম্পন্ন সমতীক্ষ্ণ, অতিতীক্ষ্ণ এবং স্বল্পতীক্ষ্ণ তিনটি প্রতিসম বিভাজন দেখানো হয়েছে।

চিত্র 6.1

অভিন্ন গাণিতিক গড় ও প্রমাণবিচ্যুতি-বিশিষ্ট অতিতীক্ষ্ণ, সমতীক্ষ্ণ এবং স্বল্পতীক্ষ্ণ তিনটি প্রতিসম পরিসংখ্যা-বিভাজন রেখা।

$b_2 = g_2 + 3$ প্রকাশনটিও অনেকসময় তীক্ষ্ণতা-মাপক হিসাবে ব্যবহৃত হয়।

$b_1$ ও $b_2$-কে যথাক্রমে পিয়ার্সনের (Pearson) $b_1$ ও $b_2$-অঙ্ক বলা হয়ে থাকে।

$b_1$ এবং $b_2$ উভয়ের ওপরই নির্ভরশীল আর একটি প্রতিবৈষম্য-মাপক হ'ল পিয়ারসনের প্রতিবৈষম্য-মাপক $P$, যার সূত্র

$$P = \frac{\sqrt{b_1}(b_2+3)}{2(5b_2 - 6b_1 - 9)} \qquad \cdots \qquad \cdots \qquad (6.23)$$

শ্রেণীবিন্যস্ত বিভাজনে এক বা একাধিক শ্রেণী অনির্দিষ্ট দৈর্ঘ্যসম্পন্ন হলে $m_4$, $s$ ইত্যাদি নির্ণয়ে অসুবিধা হয়। সেক্ষেত্রে,

$$K_2 = Q/(P_{90} - P_{10}) \qquad \cdots \qquad (6.24)$$

প্রকাশনটি ($Q$ = চতুর্থক বিচ্যুতি, $P_l$ = $l$-ক্রমিক শততমক) অনেকসময় তীক্ষ্ণতা-মাপক হিসাবে ব্যবহৃত হয়।

**উদা. 6.3** 3.9 সারণীতে প্রদত্ত রাশিতথ্যের জন্য বিভিন্ন প্রতিবৈষম্য-মাপক ও তীক্ষ্ণতা-মাপকের মান নির্ণয় করা যাক।

$$SK_1 = \frac{95{\cdot}8028 - 95{\cdot}0088}{5{\cdot}1786} = 0{\cdot}1533$$

$$SK_2 = \frac{3(95{\cdot}8028 - 95{\cdot}4500)}{5{\cdot}1786} = 0{\cdot}2044$$

$$SK_3 = \frac{99{\cdot}2283 + 92{\cdot}0387 - 2 \times 95{\cdot}4500}{99{\cdot}2283 - 92{\cdot}0387} = 0{\cdot}0510$$

$$b_1 = \frac{m_3^2}{m_2^3} = 0{\cdot}0380$$

$$g_1 = \sqrt{b_1} = 0{\cdot}1949 \text{ ($m_3$ ধনাত্মক ব'লে $g_1$-ও ধনাত্মক)}$$

$$b_2 = \frac{m_4}{m_2^2} = 3{\cdot}9327$$

$$g_2 = b_2 - 3 = 0{\cdot}9327$$

$$P = \frac{{\cdot}1949 \times 6{\cdot}9327}{2 \times 10{\cdot}4355} = {\cdot}0647.$$

আবার, $$P_{90} = 102{\cdot}95 + \frac{137{\cdot}7 - 135}{14} \times 3 = 103{\cdot}5286$$

$$P_{10} = 87{\cdot}95 + \frac{15{\cdot}3 - 8}{19} \times 3 = 89{\cdot}1026$$

এবং $Q = 3{\cdot}5948$

সুতরাং, $$K_2 = \frac{3{\cdot}5948}{103{\cdot}5286 - 89{\cdot}1026} = {\cdot}2492.$$

সবকটি প্রতিবৈষম্য-মাপকের বিচারেই বিভাজনটি সামান্য দক্ষিণায়ত প্রতিবিষম মনে হচ্ছে। এদিকে $g_2$ ও $K_2$-এর বিচারে দেখা যাচ্ছে বিভাজনটি কিছুটা

## 6.8 অনুশীলনী

6.1 বিভিন্ন ধরনের পরিঘাতের সংজ্ঞা দাও। একটি প্রদত্ত পরিসংখ্যা-বিভাজনের বিভিন্ন গড়কেন্দ্রিক পরিঘাত নির্ণয়ের কৌশল বর্ণনা কর। শার্লিয়ারের শুদ্ধি পরীক্ষার সূত্রটি কী?

6.2 প্রতিবৈষম্য কাকে বলে ? প্রমাণ কর যে, একটি প্রতিসম পরিসংখ্যা-বিভাজনের ক্ষেত্রে অযুগ্মক্রমিক যে কোন গড়কেন্দ্রিক পরিঘাতের মান শূন্য।

প্রচলিত প্রতিবৈষম্য-মাপকগুলির উল্লেখ কর এবং প্রমাণসহ এগুলির মানসীমা সম্বন্ধে আলোকপাত কর।

6.3 পরিসংখ্যা-বিভাজনের তীক্ষ্ণতা বলতে কী বোঝ ? প্রচলিত তীক্ষ্ণতা-মাপকগুলির বিবরণ দাও।

6.4 গোষ্ঠীবন্ধন ভ্রান্তি কাকে বলে ? গোষ্ঠীবন্ধন ভ্রান্তি দূর করার জন্য শেপার্ডের শুদ্ধিগুলির উল্লেখ কর। এই শুদ্ধিগুলি কোন্ কোন্ পরিস্থিতিতে প্রয়োগ করা যায় ?

6.5 $a$-কেন্দ্রিক $r$-তম পরিঘাত ${}_am'_r$ দ্বারা সূচিত হলে প্রমাণ কর যে,

$$ {}_am'_r = {}_bm'_r + \binom{r}{1}\, {}_bm'_{r-1}.d + \binom{r}{2}\, {}_bm'_{r-2}.d^2 + \cdots + d^r, $$

যেখানে $d = b - a$.

6.6 প্রমাণ কর : (i) $b_1 \geqslant 0$, (ii) $b_2 \geqslant b_1$, (iii) $b_2 \geqslant 1$, (iv) $m_{2a} \geqslant m_a^2$, (v) $b_2 - b_1 - 1 \geqslant 0$.

[ ইঙ্গিত : কোশি-শোয়াৎ´জ্‌ অসমতা-সম্পর্কটি ব্যবহার ক'রে এগুলি প্রমাণ করা যায়। (v) এর জন্য $u_i = \dfrac{x_i - \bar{x}}{s}$ এবং $v_i = \left(\dfrac{x_i - \bar{x}}{s}\right)^2 - 1$ বসাও।

বিকল্পভাবে, $X_i = x_i - \bar{x}$ হলে, $\dfrac{1}{n}\sum (AX_i^2 + BX_i + c)^2$—$A, B$ ও $C$ সম্পর্কিত এই দ্বিঘাত প্রকাশনটির (quadratic form) সুনির্দিষ্ট ধনাত্মক (positive definite) হওয়ার সর্তটি কাজে লাগিয়েও এগুলি প্রমাণ করা যেতে পারে। ]

6.7 কিছুসংখ্যক নীরেট ধাতব গোলকের ব্যাসগুলির গড় = 50 মি.মি., মধ্যমা = 48 মি.মি., $m_2 = 10$ বর্গ মি.মি. এবং $m_3 = 2$ ঘন মি.মি. দেওয়া আছে। যদি $d$ ব্যাসবিশিষ্ট একটি গোলকের ওজন হয় $5d^3$, তাহলে গোলকগুলির ওজনের গড় ও মধ্যমা নির্ণয় কর।

6.8 3.7 ও 3.8 অনুশীলনীতে প্রদত্ত রাশিতথ্য থেকে পাওয়া পরিসংখ্যা-বিভাজনের জন্য $m_3$, $m_4$ এবং বিভিন্ন প্রতিবৈষম্য-মাপক ও তীক্ষ্ণতা-মাপকগুলির মান নির্ণয় কর।

6.9 কোন পরীক্ষায় 2,350 জন পরীক্ষার্থীর ইংরেজীতে প্রাপ্ত নম্বরের

পরিসংখ্যা-বিভাজন (সমদৈর্ঘ্য শ্রেণীবিন্যাস) থেকে 38-কে কেন্দ্র ধ'রে শ্রেণীদৈর্ঘ্যের এককে বিভিন্ন পরিঘাতগুলি পাওয়া গেছে এইরকম :

$$m'_1 = 0\cdot29571, \quad m'_2 = 4\cdot8184$$

$$m'_3 = 4\cdot2592, \quad m'_4 = 71\cdot2537$$

যদি শ্রেণীদৈর্ঘ্য = 5 একক হয়, তাহলে 38-এর পরিবর্তে 50-কে কেন্দ্র ধ'রে যথার্থ এককে পরিঘাতগুলির মান নির্ণয় কর।

6.10 5.16 অনুশীলনীতে প্রদত্ত 5টি ভুল নম্বরের ভিত্তিতে নির্ণীত $m_3$ ও $m_4$-এর মান যথাক্রমে 112·62 এবং 3129·92 হলে এগুলির সঠিক মান নির্ণয় কর।

## 6.9 নির্দেশিকা

1. Cook, L. H. L. *Statistical Problems and How to Solve Them.* Barnes and Noble, 1971.

2. Goon A. M., Gupta, M. K., & Dasgupta B. *Fundamentals of Statistics, Vol I.* World Press, 1975.

3. Mounsey, J. *Introduction to Statistical Calculations.* English University Press, 1952.

4. Kendall, M. G. & Stuart, A. *Advanced Theory of Statistics, Vol. I.* Charles Griffin, 1960.

5. Mills, F. C. *Statistical Methods.* H. Holt, 1955.

6. Yule, G. U., & Kendall M. G. *Introduction to the Theory of Statistics.* Charles Griffin, 1953.

# সম্ভাবনাতত্ত্বের প্রাথমিক আলোচনা 7
## ( Basic ideas of probability theory )

### 7.1 সম্ভাবনার স্বরূপ (meaning of probability) :

বর্তমান পরিচ্ছেদে সম্ভাবনার সংজ্ঞা ও তার কিছু কিছু বৈশিষ্ট্য সম্পর্কে সংক্ষেপে কয়েকটি বিষয় আলোচনা করব।

সম্ভাবনার সংজ্ঞা নির্দেশ করতে হলে সর্বদাই একটি **সম্ভাবনাভিত্তিক পরীক্ষণের** (random experiment) কথা আমাদের মনে রাখতে হবে। যে ক্রিয়ার অনুষ্ঠানের সূত্রে কোন ফলাফলের (outcomes) অবেক্ষণ (observation) সম্ভব, ব্যাপক অর্থে তাকেই আমরা পরীক্ষণ বলব। কিন্তু যে পরীক্ষা সম্পন্ন হবার আগেই তার কী ফল ঘটবে তা নিশ্চিতভাবে জানা যায় সম্ভাবনাতত্ত্বের প্রসঙ্গে তা মোটেই বিবেচ্য নয়। সম্ভাবনাভিত্তিক পরীক্ষণ হচ্ছে শুধু তেমন পরীক্ষণ যার অবেক্ষণযোগ্য ফলগুলি (observable outcomes) কী হতে পারে তা জানা থাকলেও পরীক্ষণের বিশেষ কোন অনুষ্ঠানে কোন্ ফলটি বাস্তবিক ঘটবে তা আগেই জানা বা অভ্রান্তভাবে অনুমান করা যায় না। যেমন, একটি মুদ্রা নিক্ষিপ্ত হলে তার দুটি পার্শ্বের একটি দেখা যায় মুদ্রাটির ওপরে। এখানে নিক্ষেপণ কাজটি হচ্ছে একটি পরীক্ষণ। এর অবেক্ষণযোগ্য ফল হচ্ছে দুটি : কারণ, মুদ্রাটি নিক্ষিপ্ত হবার পর তার ওপরের দিকে 'অশোকচক্র' ( যাকে আমরা মুদ্রার 'সম্মুখপার্শ্ব' বলব ) বা 'ধান্যশীর্ষ'—( যাকে আমরা মুদ্রার 'পশ্চাৎপার্শ্ব' বলব ) চিহ্নিত পার্শ্বদুটির যে কোন একটি দেখা যেতে পারে। এটি একটি সম্ভাবনা-ভিত্তিক পরীক্ষণ। কারণ, কোন্ পার্শ্বটি বাস্তবিক দেখা যাবে তা মুদ্রা উৎক্ষেপণের আগে নির্ভুলভাবে অনুমান করা যায় না।

সম্ভাবনাভিত্তিক পরীক্ষণের অনুষ্ঠানে অবেক্ষণযোগ্য কোন ফলকে বলা হয় একটি **'ঘটনা'** (event) বা **'সম্ভাবনানির্ভর ঘটনা'** (random event). একটি লুডো খেলার ছক্কা নিক্ষিপ্ত হলে সেটি যখন স্থির হয়ে দাঁড়াবে তখন তার সবচেয়ে ওপরের প্রান্তে 1, 2, 3, 4, 5 বা 6 সংখ্যা-নির্দেশক চিহ্নের যে কোন একটিকে দেখা যাবে। এখানে ছক্কা নিক্ষেপণ হচ্ছে একটি সম্ভাবনাভিত্তিক পরীক্ষণ এবং এই ছটি সংখ্যার যে কোন একটি নির্দেশক চিহ্নযুক্ত পার্শ্বটি ছক্কাটির ওপরে থাকাই হচ্ছে এক একটি ঘটনা। এখানে উল্লেখযোগ্য যে,

পরীক্ষণের যে ফল একাধিক বিভিন্ন ফলের সমাহাররূপে অবেক্ষণযোগ্য নয় তাকে বলা হয় একটি **'মৌলিক ঘটনা'** (elementary event). যেমন, ছক্কার ওপরে 3 ( বা 4 বা অন্য যে কোন সংখ্যা )-নির্দেশক চিহ্ন দৃষ্ট হওয়া ব্যাপারটি একটি মৌলিক ঘটনা। কিন্তু ছক্কার ওপরে 'যুগ্ম সংখ্যা নির্দেশক চিহ্ন' দৃষ্ট হওয়া হচ্ছে একটি ঘটনা ( মৌলিক ঘটনা নয় )। সাধারণভাবে একটি ঘটনা কয়েকটি মৌলিক ঘটনার সমবায়ে গঠিত হতে পারে। যেমন, ছক্কার ওপর 2 বা 4 বা 6 দৃষ্ট হওয়ার মৌলিক ঘটনার যে কোন একটি ঘটলেই 'যুগ্ম সংখ্যা নির্দেশক চিহ্ন' দৃষ্ট হওয়ার ঘটনাটি ঘটবে। কাজেই বলা যায় যে ঘটনা হচ্ছে পরীক্ষণ ফলের একটি গুচ্ছ (set) যা আরও সরলতররূপে অবেক্ষণযোগ্য (further decomposable into simpler elements). বাস্তবিক, ঐ গুচ্ছের যে কোন একটি ঘটলেই ঐ গুচ্ছনির্দেশিত ঘটনাটি ঘটেছে ব'লে স্বীকার করা যায়। কিন্তু কোন মৌলিক ঘটনা অধিকতর সরলরূপে অবেক্ষণযোগ্য নয়। কোন পরীক্ষণের সঙ্গে সংশ্লিষ্ট সব কটি মৌলিক ঘটনার সমবায়ে যে গুচ্ছ গঠিত হয়, তাকে বলে পরীক্ষণটির **নমুনা দেশ** (sample space). এখন, সম্ভাবনাভিত্তিক কোন পরীক্ষণ-ক্রিয়া সম্পন্ন হলে তাতে কোন বিশেষ ঘটনা আদৌ ঘটবে কিনা তা নিশ্চিতভাবে জানা যায় না ব'লেই অনেকসময়ই আমাদের জানতে কৌতূহল হয়, ঐ ঘটনাটি ঘটবার **'সম্ভাবনা'** (probability) কত? যেমন, উৎক্ষিপ্ত মুদ্রার ওপরে 'সম্মুখপার্শ্ব' দেখা যাবার নিশ্চয়তা নেই ব'লেই জানতে ইচ্ছে করে, এরকম ঘটনার সম্ভাবনা কতটুকু। এই যে 'সম্ভাবনা' কথাটি বলা হচ্ছে এর প্রকৃত সংজ্ঞা কী? এই প্রসঙ্গে আমরা এখন সম্ভাবনার **'পুরাতনী'** (classical) সংজ্ঞা নিয়ে আলোচনা করব।

### 7.2 সম্ভাবনার পুরাতনী সংজ্ঞা (classical definition of probability) :

উনবিংশ শতাব্দীর প্রখ্যাত গাণিতিক **লাপ্লাস** (Laplace), **বেরণুলি** (Bernoulli) এবং তাদের মতানুসারী কতিপয় মনীষীর আলোচনার সূত্রেই সম্ভাবনার **পুরাতণী তত্ত্বটি** (classical theory) গড়ে উঠেছিল। এই তত্ত্বে প্রথমেই ধ'রে নেওয়া হয় যে, সম্ভাবনাভিত্তিক পরীক্ষণটি হচ্ছে **সুষম প্রকৃতি-বিশিষ্ট** (symmetric) এবং এর মোট মৌলিক ঘটনার সংখ্যা সসীম। পরীক্ষণটি সুষম বলতে মোটামুটি আমরা যা বুঝি তা হচ্ছে এই যে, এটি সম্ভাব্য মৌলিক ঘটনাগুলির মধ্যে কোনটির অনুকূলে বা প্রতিকূলেই কোন পক্ষপাত দর্শাবে না।

এই পক্ষপাতহীনতার লক্ষণ হচ্ছে এই যে পরীক্ষণটি যদি বহুবার অনুষ্ঠিত হয় তবে কোন ফলই অপর কোন ফলের চেয়ে লক্ষনীয়ভাবে অধিকতর সংখ্যায় সংঘটিত হবে না। সংঘটনের সংখ্যাসাম্যের নিরিখ বাদ দিয়েও পরীক্ষণটির গুণলক্ষণ ও চারিত্রিক বৈশিষ্ট্য এবং ঘটনাগুলির গঠনই এমন হতে হবে যে সাধারণবুদ্ধিতে কখনই যেন এমন মনে না হয় যে, পরীক্ষণটি কোন এক বা একাধিক মৌলিক ঘটনার অনুকূলে তার ফল দর্শাতে পারে। এক্ষেত্রে এই মৌলিক ঘটনাগুলির প্রত্যেকটিকে **সমসম্ভব** (equally likely) ব'লে উল্লেখ করা হয়ে থাকে। উদাহরণতঃ, একটি ছক্কার গঠন যদি স্বাভাবিক হয় তবে এর প্রত্যেকটি পার্শ্বই সমান আকৃতি, মসৃণতা এবং ওজনবিশিষ্ট হবে। ফলে, এটি গড়িয়ে দিলে তার ছটি প্রান্তের কোন একটি বিশেষ প্রান্ত (অপর প্রান্তগুলির পরিবর্তে) ছক্কাটির ওপরে দেখা যাওয়ার সম্ভাবনা বেশি, এমন আশা করার সঙ্গত কারণ থাকে না।

যাই হোক, আমরা ধ'রে নেব যে পরীক্ষণটি অনুষ্ঠিত হলে মোট $N$ (সসীম) সংখ্যক বিভিন্ন মৌলিক ঘটনা ঘটতে পারে এবং তারা সকলেই সমসম্ভব। পরিভাষা অনুযায়ী বলা হয় যে পরীক্ষণটি এমন যে এর মোট সমসম্ভব পরিস্থিতির (equally likely cases) সংখ্যা $N$. এখন মনে করা যাক যে, এই পরীক্ষণের সঙ্গে সংশ্লিষ্ট একটি ঘটনা সম্পর্কে আমরা আগ্রহী যাকে আমরা সংকেতচিহ্ন $A$ দ্বারা নির্দেশ করব। আমরা জানতে চাই পরীক্ষণটি অনুষ্ঠিত হলে $A$ ঘটনাটির সংঘটিত হবার সম্ভাবনা কত? ধরা যাক $A$ হচ্ছে মোট $N(A)$ সংখ্যক মৌলিক ঘটনাবলীর একটি গুচ্ছ। অর্থাৎ যখনই $N(A)$ সংখ্যক নির্দিষ্ট অবেক্ষণ-যোগ্য মৌলিক ঘটনার যে কোন একটি ঘটবে তখন $A$ ঘটনাটি ঘটেছে ব'লে স্বীকার্য। এক্ষেত্রে পরিভাষা অনুযায়ী বলা হবে যে মোট $N$ সংখ্যক সমসম্ভব পরিস্থিতির মধ্যে $N(A)$ সংখ্যক পরিস্থিতি হচ্ছে $A$ ঘটনার অনুকূল (favourable to the event $A$). এখানে অবশ্যই $N(A) \leqslant N$. তাহলে সম্ভাবনার পুরাতনী সংজ্ঞা অনুযায়ী $A$ ঘটনার সম্ভাবনা হচ্ছে $\frac{N(A)}{N}$. এই সম্ভাবনাকে আমরা $P(A)$ —এই সংকেতসূত্রে প্রকাশ করব; অর্থাৎ

$$P(A)=\frac{N(A)}{N} \qquad \cdots \qquad \cdots \quad (7.1)$$

হচ্ছে $A$ ঘটনার সম্ভাবনা।

## 7.3 কয়েকটি উদাহরণ :

সম্ভাবনার পুরাতনী সংজ্ঞা গ্রহণ ক'রে তার ভিত্তিতে এখন আমরা কয়েকটি ঘটনার সম্ভাবনার মান নির্ণয় করব।

কোন পরীক্ষণ-এর মোট সমসম্ভব পরিস্থিতিসংখ্যা $N$ হলে সংশ্লিষ্ট মৌলিক ঘটনাগুলিকে $\omega_1, \ldots, \omega_N$ এবং তাদের কয়েকটির সমবায়ে গঠিত ঘটনাকে সাধারণভাবে $A = \{\omega_{i_1}, \ldots \omega_{i_k}\}$ সংকেতচিহ্নের সাহায্যে প্রকাশ করব। এখানে $i_1, \ldots i_k$ হচ্ছে প্রথম $N$টি অখণ্ডসংখ্যার যে কোন $k$টি বিভিন্ন সংখ্যাবিশেষ।

**উদা. 7.1** তিনটি মুদ্রা একসঙ্গে নিক্ষিপ্ত হলে দুটিতে মুদ্রার 'সম্মুখপার্শ্ব' ওপরে দৃষ্ট হওয়ার সম্ভাবনা কত ?

$H$ এবং $T$ যদি যথাক্রমে মুদ্রার সম্মুখ ও পশ্চাৎপার্শ্ব দৃষ্ট হওয়ার ঘটনা নির্দেশ করে, তবে এক্ষেত্রে তিনটি মুদ্রা নিক্ষেপণের পরক্ষীণ-সংশ্লিষ্ট মৌলিক ঘটনাগুলি হচ্ছে

$\omega_1 : HHH, \omega_2 : HHT, \omega_3 : HTH, \omega_4 : HTT,$

$\omega_5 : THH, \omega_6 : THT, \omega_7 : TTH$ এবং $\omega_8 : TTT.$

সম্ভাবনার পুরাতনী তত্ত্বানুযায়ী এই আটটি মৌলিক ঘটনার প্রত্যেকটিকে সমসম্ভব বলে স্বীকার করা হয় এবং এক্ষেত্রে পরীক্ষণটির মোট সমসম্ভব পরিস্থিতি সংখ্যা হচ্ছে $N = 8$. আমাদের বিবেচ্য ঘটনা $A$ হচ্ছে সেই ঘটনা যা ঘটবে যদি দুবার মুদ্রায় সম্মুখপার্শ্ব দেখা যায় অর্থাৎ যদি $\omega_2 : HHT, \omega_3 : HTH,$ বা $\omega_5 : THH$-এর যে কোন একটি মৌলিক ঘটনা ঘটে। অর্থাৎ $A = \{\omega_2, \omega_3, \omega_5\}$. তাহলে $A$ ঘটনার অনুকূল পরিস্থিতিসংখ্যা হচ্ছে $N(A) = 3$. সুতরাং $A$ ঘটনাটির নির্ণেয় সম্ভাবনা $P(A) = \frac{3}{8}$.

**উদা. 7.2** যদি $TOWEL$ শব্দটিতে ব্যবহৃত অক্ষরগুলিকে সমসম্ভব উপায়ে সাজানো যায় তবে $O$ এবং $E$-এর মাঝখানে অন্য দুটি অক্ষর থাকবার সম্ভাবনা কত ?

এখানে সম্ভাবনানির্ভর পরীক্ষণটি হচ্ছে $T, O, W, E$ ও $L$ এই পাঁচটি অক্ষরকে পরপর এমনভাবে সাজানো, যাতে প্রত্যেকটি পৃথক্ বিন্যাস পরিদৃষ্ট হবার সম্ভাবনা সমান থাকে। যে পাঁচটি অবস্থিতিতে এই অক্ষরগুলিকে বসাতে হবে তাদেরকে 1, 2, 3, 4 এবং 5 নম্বর দিয়ে চিহ্নিত করলে এদের থেকে দুটিকে $\binom{5}{2} = 10$ সংখ্যক উপায়ে বেছে নিয়ে ঐ দুটিতে $O$ এবং $E$ অক্ষর-দুটিকে সন্নিবেশিত করা যায়। এই দশটি উপায়ে নির্বাচিত প্রত্যেকটি বিন্যাসকে এই

পরীক্ষণ-সংশ্লিষ্ট এক একটি মৌলিক ঘটনা বলা হলে এক্ষেত্রে মোট সমসম্ভব পরিস্থিতি সংখ্যা হবে 10. এখন আলোচ্য ঘটনা $A$ ঘটবে যদি $O$ এবং $E$-কে 1 এবং 4 অথবা 2 এবং 5 চিহ্নিত অবস্থিতিতে সন্নিবেশিত করা হয়। তাহলে $A$ ঘটনার অনুকূল পরিস্থিতি সংখ্যা 2. সুতরাং $A$ ঘটনার সম্ভাবনা $P(A) = \frac{2}{10} = \frac{1}{5}$.

**উদা. 7.3** ছটি ছেলেমেয়ে যদি বৃত্তাকারে দাঁড়ায় তবে তাদের মধ্যে বিশেষ দুজনের মাঝখানে ঠিক তিনজন অন্য ছেলেমেয়ে থাকবার সম্ভাবনা নির্ণয় কর। [ এখানে ঘড়ির কাঁটার দিক (clockwise direction) অনুযায়ী হিসেব ক'রে ঠিক করতে হবে কে কার পরে দাঁড়াচ্ছে। ]

এখানে সম্ভাবনাশ্রয়ী পরীক্ষণটি হচ্ছে এই যে, ছটি ছেলেমেয়ে বৃত্তাকারে দাঁড়াবার সময় তাদের পারস্পরিক স্থান এমনভাবে বেছে নেবে যেন প্রত্যেকটি নির্বাচন সমসম্ভব হয়। এখন ছটি স্থানকে 1, 2, 3, 4, 5 এবং 6 নম্বর দিয়ে চিহ্নিত ক'রে তাদের থেকে দুটি স্থান $\binom{6}{2} = 15$ রকম বিভিন্ন উপায়ে বেছে নিয়ে ঐ স্থান-দুটিতে ঐ বিশেষ ছেলেমেয়ে-দুটিকে দাঁড় করানো যায়। এভাবে যে 15 রকম বিভিন্ন বিন্যাস পাওয়া যায় ঐগুলিকে এই পরীক্ষণ-সংশ্লিষ্ট মৌলিক ঘটনা বলা যাক। তাহলে পরীক্ষণের মোট সমসম্ভব পরিস্থিতিসংখ্যা 15. এখন উদাহরণে উল্লিখিত ঘটনাটি ঘটবে যদি (1, 5), (2, 6), (3, 1), (4, 2), (5, 3) এবং (6, 4)-এর মধ্যে যে কোন একটি সংখ্যাদ্বৈতের প্রথম ও দ্বিতীয় সংখ্যক স্থান-দুটিতে ঐ বিশেষ ছেলেমেয়ে-দুটি দাঁড়ায়। তাহলে এই ঘটনার মোট অনুকূল পরিস্থিতি হচ্ছে 6টি। সুতরাং নির্ণেয় সম্ভাবনা $\frac{6}{15} = \frac{2}{5}$.

**উদা. 7.4** $1, 2, \ldots x-1, x$ রাশিসমূহের এক একটি দ্বারা চিহ্নিত, কিন্তু অন্য সর্বপ্রকারে অভিন্ন, $x$-সংখ্যক টিকিট থেকে যদি তিনটি টিকিট সমসম্ভাবনা সহকারে বেছে নেওয়া হয়, তাহলে ঐ তিনটি টিকিটে চিহ্নিত সংখ্যাত্রয় সমান্তরশ্রেণী গঠন করার সম্ভাবনা কত ?

এখানে পরীক্ষণটি হচ্ছে $x$ সংখ্যক টিকিট থেকে তিনটি টিকিট বেছে নেওয়া, যাতে প্রত্যেকটি নির্বাচন সমসম্ভব হয়। এখানে পরীক্ষণটি সুষম। কারণ, টিকিটগুলির একমাত্র তফাৎ হচ্ছে এই যে তাদের গায়ে উৎকীর্ণ সংখ্যাগুলি পৃথক্। এটা অবশ্য ধ'রে নেওয়া হবে যে টিকিটগুলি এমনভাবে তোলা হবে যেন তাদের গায়ে লেখা সংখ্যাগুলি চোখে না পড়ে। এখন এই নির্বাচন $\binom{x}{3}$ সংখ্যক বিভিন্ন

উপায়ে করা যায়। এই প্রত্যেকটি বিভিন্ন নির্বাচনকে এই পরীক্ষণের মৌলিক ঘটনা বলা যায়। কাজেই এক্ষেত্রে মোট সমসম্ভব পরিস্থিতিসংখ্যা হচ্ছে $\binom{x}{3}$.

এখন $x$-কে একটি যুগ্মরাশি $= 2n$ ধ'রে সমস্যাটির সমাধান কী হয় দেখা যাক। নির্বাচনে প্রাপ্ত টিকিটের সংখ্যা তিনটি $w_1$, $w_2$, $w_3$ হলে যদি বলা হয় যে, সংঘটিত মৌলিক ঘটনাটি হচ্ছে, $\{w_1, w_2, w_3\}$, তাহলে উদাহরণে উল্লিখিত ঘটনা $A$-এর অনুকূল পরিস্থিতিগুলি হচ্ছে নিম্নরূপ :—

$\{1, 2, 3\}$, $\{1, 3, 5\}$, $\{1, 4, 7\}$, ..., $\{1, n, 2n-1\}$ ;
$\{2, 3, 4\}$, $\{2, 4, 6\}$, $\{2, 5, 8\}$, ..., $\{2, n+1, 2n\}$ ;
$\{3, 4, 5\}$, $\{3, 5, 7\}$, $\{3, 6, 9\}$, ..., $\{3, n+1, 2n-2\}$ ;
... ... ...
... ... ...
$\{2n-3, 2n-2, 2n-1\}$ ;

এবং $\{2n-2, 2n-1, 2n\}$.

তাহলে $A$ ঘটনার মোট অনুকূল পরিস্থিতিসংখ্যা হচ্ছে

$$2\{1+2+\cdots+(n-2)+(n-1)\} = 2 \times \frac{n(n-1)}{2} = n(n-1).$$

এখন, $\binom{x}{3} = \binom{2n}{3} = \frac{2n(2n-1)(2n-2)}{6}$.

সুতরাং ঘটনাটির নির্ণেয় সম্ভাবনা হচ্ছে $\frac{3}{2(2n-1)}$.

অনুরূপভাবে দেখানো যায় যে, $x$ একটি অযুগ্মরাশি $= 2n+1$ হলে সম্ভাবনার মান দাঁড়াবে $\frac{3n}{4n^2-1}$.

**উদা. 7.5** খুব ভালোভাবে মেশানো পুরো বাহান্নটি তাসের একটি প্যাকেট থেকে সমান সম্ভাবনা আরোপ ক'রে তিনটি তাস বেছে নিলে সেগুলির প্রতিটিই টেক্কা হবার সম্ভাবনা কত ?

এখানে পরীক্ষণটি হচ্ছে বাহান্নটি তাস থেকে তিনটি তাস সমসম্ভাবনা-সহকারে বেছে নেওয়া। তাসগুলির মধ্যে আকারে ও ওজনে কোন তফাৎ নেই। তাই এদের যদি মান ও বর্ণ না দেখে নেওয়া হয় তাহলে স্পষ্টতঃই পরীক্ষণটি সুষম ব'লে মেনে নিতে কোন আপত্তি নেই। এখন 52টি তাস থেকে 3টি তাস $\binom{52}{3}$ সংখ্যক বিভিন্ন উপায়ে বেছে নেওয়া যায়। এই প্রত্যেকটি

নির্বাচন সমসম্ভব এবং এরাই এক একটি মৌলিক ঘটনা। তাহলে মোট সমসম্ভব পরিস্থিতিসংখ্যা হচ্ছে $\binom{52}{3}$. প্যাকেটটিতে মোট চারটি টেক্কা রয়েছে। তাদের থেকে তিনটি টেক্কা মোট $\binom{4}{3}$ সংখ্যক উপায়ে বেছে নেওয়া যায়। কাজেই উল্লিখিত ঘটনাটিকে সংকেত চিহ্ন $A$ দ্বারা সূচিত করলে এর অনুকূল পরিস্থিতিসংখ্যা হচ্ছে $\binom{4}{3}$. সুতরাং ঘটনাটির সম্ভাবনা হচ্ছে $\dfrac{\binom{4}{3}}{\binom{52}{3}}=\dfrac{1}{5525}$.

**উদা. 7.6** খুব ভালোভাবে মেশানো পুরো বাহান্নটি তাসের একটি প্যাকেট থেকে সমান সম্ভাবনা আরোপ ক'রে চারটি তাস বেছে নিলে তাদের মধ্যে দুটি টেক্কা পাবার সম্ভাবনা নির্ণয় কর।

উদা. 7.5-এ বর্ণিত যুক্তি অনুযায়ী এখানে মোট সমসম্ভব পরিস্থিতি-সংখ্যা হচ্ছে $\binom{52}{4}$. আবার, উল্লিখিত ঘটনাটির অনুকূল পরিস্থিতিসংখ্যা হচ্ছে $\binom{4}{2}\binom{48}{2}$; কারণ চারটি টেক্কা থেকে দুটি টেক্কা $\binom{4}{2}$ সংখ্যক বিভিন্ন উপায়ে বেছে নেওয়া যায় এবং সঙ্গে সঙ্গে অন্য দুটি তাস টেক্কা ছাড়া বাকী 48টি তাস থেকে $\binom{48}{2}$ সংখ্যক উপায়ে নেওয়া যায়। কাজেই চারটি তাসের মধ্যে দুটি টেক্কা ও অন্য দুটি টেক্কা ছাড়া তাস মোট $\binom{4}{2}\times\binom{48}{2}$ সংখ্যক উপায়ে বেছে নেওয়া যায়। কাজেই নির্ণেয় সম্ভাবনা হ'ল $\dfrac{\binom{4}{2}\binom{48}{2}}{\binom{52}{4}}$,

**উদা. 7.7** মনে কর, সম-আকৃতিবিশিষ্ট তিনটি বাক্সের প্রত্যেকটিতে দুটি ক'রে প্রকোষ্ঠ রয়েছে এবং প্রত্যেক প্রকোষ্ঠে একটি ক'রে বল আছে। যদি একটি বাক্সের দুটি বলই সাদা, আর একটি বাক্সের দুটি বলই কালো এবং অপরটিতে একটি সাদা ও একটি কালো হয়, তাহলে সমসম্ভব উপায়ে একটি বাক্স বেছে নিয়ে তার একটি প্রকোষ্ঠের বলটিকে যদি সাদা দেখা যায়, তবে অপরটি কালো হওয়ার সম্ভাবনা কত?

এখানে পরীক্ষণটি হচ্ছে তিনটি অভিন্ন আকৃতির বাক্সের একটিকে বেছে নিয়ে তার যে কোন একটি প্রকোষ্ঠের বলটিকে দেখা। যেহেতু বাক্স তিনটিকে আপাত-দৃষ্টিতে পৃথক্ মনে করার কারণ নেই, কাজেই ধরা যেতে পারে যে পরীক্ষণটি সুষম। এখন সাদা বল-ভর্তি মোট 3টি প্রকোষ্ঠ রয়েছে। কাজেই পরীক্ষণের মোট সমসম্ভব পরিস্থিতিসংখ্যা হচ্ছে 3 ; কারণ, এদেরই একটি প্রকোষ্ঠ বেছে নিয়ে তার মধ্যে সাদা বল পাওয়া গেছে। এখন, এদের মধ্যে একটিমাত্র এমন প্রকোষ্ঠ রয়েছে যে, যে বাক্সের মধ্যে এটি আছে সেই বাক্সের অপর প্রকোষ্ঠে যে বলটি আছে তার রঙ কালো। কাজেই উদাহরণে বর্ণিত ঘটনাটির অনুকূল পরিস্থিতিসংখ্যা 1. কাজেই নির্ণেয় সম্ভাবনার মান $\frac{1}{3}$.

## 7.4 কয়েকটি সংজ্ঞা ঃ

$A$ ও $B$ যদি দুটি ঘটনা নির্দেশ করে, তবে $A$ ও $B$-এর যুগপৎ সংঘটনের ঘটনাকে আমরা $A \cap B$ ( অথবা $AB$) সংকেত চিহ্নের মাধ্যমে প্রকাশ করব। এ জাতীয় ঘটনাকে অনেক সময় **মিশ্র-ঘটনা** (compound event) বলা হয়। ছক্কা নিক্ষেপণের পরীক্ষণে $w_i$ ($i=1, 2..., 6$) যদি $i$-সংখ্যা নির্দেশক চিহ্নের আবির্ভাব বোঝায় তাহলে $A=\{w_2, w_4, w_6\}$ এবং $B=\{w_3, w_6\}$ হচ্ছে যথাক্রমে যুগ্ম-সংখ্যক চিহ্ন এবং 3 এর গুণনীয়কবিশিষ্ট চিহ্ন দৃষ্ট হওয়ার ঘটনা। কাজেই $A \cap B=\{w_6\}$ বোঝাবে 6-সূচক চিহ্নের আবির্ভাব।

প্রত্যেক সম্ভাবনাত্মক পরীক্ষণের সঙ্গেই সাধারণতঃ দুটি ঘটনা সর্বদাই জড়িত রয়েছে বলে সম্ভাবনা শাস্ত্রে ধ'রে নেওয়া হয়। তাদের একটিকে বলে **নিশ্চিত ঘটনা** (sure event) এবং অপরটিকে বলে **অসম্ভব ঘটনা** (impossible event). এমন একটি ঘটনা আছে পরীক্ষণ কার্যটি সমাধা হলেই আবশ্যিকভাবে যা ঘটতে দেখা যাবে। যদি একটি পরীক্ষণ $\varepsilon$ সাধিত হলে সর্বমোট সম্ভাব্য পরিস্থিতি হয় $w_1, w_2, ..., w_N$ এবং $\Omega=\{w_1, ..., w_i..., w_N\}$ নির্দেশ করে তাদের সবগুলির একত্র গৃহীত গুচ্ছ, তাহলে $\Omega$ হচ্ছে সেই ঘটনা যা ঘটবে যখনই পরীক্ষণের ফলস্বরূপে $w_1, ...., w_N$-এর যে কোন একটি মৌলিক ঘটনাকে ঘটতে দেখা যাবে। এখন পরীক্ষণটির গঠন-প্রকৃতিই এমন যে, যখনই পরীক্ষণটি সম্পন্ন হবে তখনই $w_1, ... w_N$-এর মধ্যে অন্ততঃ একটিকে ঘটতে দেখা যাবেই। কাজেই পরীক্ষণটি সম্পন্ন হলেই $\Omega$ ঘটনাটি ঘটবেই, অর্থাৎ ঐ বিশেষ পরীক্ষণের সঙ্গে সংশ্লিষ্ট $\Omega$ ঘটনাটি একটি নিশ্চিত ঘটনা। স্পষ্টতঃই এই নিশ্চিত ঘটনাটির সম্ভাবনার মান

হবে 1. কারণ, $\Omega$ ঘটনাটির অনুকূল পরিস্থিতিসংখ্যা এবং পরীক্ষণ-সংশ্লিষ্ট মোট সমসম্ভব পরিস্থিতিসংখ্যা উভয়ই $N$. কাজেই, সম্ভাবনার পুরাতনী সংজ্ঞানুসারে,

$$P(\Omega) = \frac{N}{N} = 1. \qquad \cdots \quad (7.2)$$

অবশ্য বিপরীতক্রমে এটা সব সময়েই বলা যাবে না যে, যদি কোন ঘটনা $A$-এর সম্ভাবনার মান 1 হয়, তাহলে $A$ ঘটনাটি একটি নিশ্চিত ঘটনা হবেই। একটি উদাহরণ নেওয়া যাক। একটি মুদ্রা উৎক্ষিপ্ত হলে ধরা যেতে পারে যে তাতে তিনটি পৃথক্ ঘটনার সংঘটন সম্ভব। সেগুলি হচ্ছে মুদ্রার ওপরে (1) 'সম্মুখপার্শ্ব' দৃষ্ট হওয়া ($H$), (2) 'পশ্চাৎপার্শ্ব' দৃষ্ট হওয়া ($T$) এবং (3) মুদ্রাটি তার প্রান্তভাগের ওপর দণ্ডায়মান থাকা ( বলা যাক $E$)। এখানে নমুনাদেশকে লেখা যেতে পারে $\Omega = (H,\ T,\ E)$ এবং এটি হচ্ছে একটি নিশ্চিত ঘটনা। কিন্তু সাধারণতঃ উৎক্ষিপ্ত মুদ্রাটি তার প্রান্তভাগের ওপর দাঁড়িয়ে থাকার ঘটনা এতই কদাচিৎ ঘটতে পারে যে এর সম্ভাবনাকে নগণ্য ধরা যেতে পারে। ফলে উৎক্ষিপ্ত মুদ্রায় 'সম্মুখ' অথবা 'পশ্চাৎ' পার্শ্বের একটি দৃষ্ট হওয়ার ঘটনার সম্ভাবনার মান 1 ব'লে ধরা হয় অর্থাৎ এটা মেনে নেওয়া হয় যে $P(H,\ T) = 1$ এবং $P(E) = 0$. কিন্তু এক্ষেত্রে $(H,\ T)$ একটি নিশ্চিত ঘটনা নয়, কারণ $E$ ঘটনার সংঘটনকে ধারণার বাইরে রাখতেই হবে এমন নয়, যদিও তার সম্ভাবনাকে ধর্তব্য নয় ব'লে মনে করা যায়।

নিশ্চিত ঘটনার মত আরও একটি ঘটনা প্রত্যেক সম্ভাবনাভিত্তিক পরীক্ষণের সঙ্গে সর্বদা জড়িত আছে ব'লে ধরা হয়। একে বলে **অসম্ভব ঘটনা** (impossible event). এটি হচ্ছে সেই ঘটনা আলোচ্য পরীক্ষণটি সম্পন্ন হলে যা কখনই ঘটতে পারে না। একে আমরা $\phi$ সংকেত চিহ্নের সাহায্যে নির্দেশ করব। এর বৈশিষ্ট্য হচ্ছে এই যে, আলোচ্য পরীক্ষণ-সংশ্লিষ্ট কোন মৌলিক ঘটনা ঘটলেই এটি ঘটবে না এবং মৌলিক ঘটনাগুলির কোন গুচ্ছের মাধ্যমেই এই ঘটনাটি ঘটতে দেখা যাবে না। ফলে, $\phi$ ঘটনার অনুকূল পরিস্থিতিসংখ্যা হবে শূন্য এবং সম্ভাবনার পুরাতনী সংজ্ঞানুযায়ী অসম্ভব ঘটনা $\phi$-এর সম্ভাবনার মানও হবে শূন্য।

$$\text{অর্থাৎ} \quad P(\phi) = 0. \qquad \cdots \quad (7.3)$$

উদাহরণস্বরূপ বলা যায় যে, দুটো লুডো খেলার ছক্কা একত্র নিক্ষিপ্ত হলে তাদের উভয়ের ওপর দৃষ্ট চিহ্ন অনুযায়ী সংখ্যা-দুটির সমষ্টি 1 হওয়ার ব্যাপারটি একটি অসম্ভব ঘটনা, কারণ প্রতিটি ছক্কার ওপর চিহ্ন অনুযায়ী সংখ্যা হচ্ছে

1, 2, 3, 4, 5 এবং 6। এবং এই ঘটনার সম্ভাবনা হচ্ছে শূন্য। উল্লেখ্য যে, কোন ঘটনার সম্ভাবনা শূন্য হলেই তা অসম্ভব ঘটনা হবেই এমন কোন কথা নেই। যেমন একটি উৎক্ষিপ্ত মুদ্রা তার ধারের ওপর খাড়াভাবে দাঁড়াবার ঘটনাটির সম্ভাবনা প্রচলিত রীতি অনুযায়ী যদিও শূন্য, কিন্তু এটি একটি অসম্ভব ঘটনা নয়।

যে কোন দুটি ঘটনা $A$ এবং $B$-এর যুগপৎ সংঘটনের ঘটনা $A \cap B$ যদি একটি অসম্ভব ঘটনা হয়, তবে $A$ এবং $B$-কে **পরস্পর ব্যতিরেকী** (mutually exclusive) ঘটনা বলা হয়। যেমন একটি লুডো খেলার ছক্কায় যুগ্মসংখ্যা-সূচক চিহ্নের আবির্ভাবকে $A$ এবং 5-এর অখণ্ড গুণনীয়ক নির্দেশক চিহ্নের উপস্থিতিকে $B$ বলা হলে, $A \cap B$ হবে অসম্ভব ঘটনা $\phi$ এবং এক্ষেত্রে $A$ ও $B$ হচ্ছে পরস্পর-ব্যতিরেকী এবং ফলে $P(A \cap B) = 0$.

যদি $n$ সংখ্যক ঘটনা $A_1, A_2, \ldots, A_i, \ldots, A_n$ এরূপ সম্পর্কযুক্ত হয় যে, তাদের মধ্যে যে কোন একটি ঘটলে বাকী ঘটনাগুলির একটিও ঘটতে পারে না, অর্থাৎ যদি তারা এমন হয় যে, প্রত্যেক $i, j = 1, \ldots, n$ $(i \neq j)$ এর জন্যে $A_i$ ও $A_j$ পরস্পর ব্যতিরেকী হয়, তাহলে এই ঘটনাগুলিকে যৌথভাবে পরস্পর ব্যতিরেকী বলা হয়। এস্থলে, প্রত্যেক জোড়া ঘটনা $A_i$ ও $A_j$ পরস্পর ব্যতিরেকী ; অর্থাৎ তারা যুগ্মভাবে পরস্পর ব্যতিরেকী।

যদি $A$ ও $B$ দুটি ঘটনা হয়, তবে $A \cup B$ সংকেত চিহ্ন ব্যবহার ক'রে আমরা নির্দেশ করব সেই ঘটনা যা ঘটবে যদি এই ঘটনা-দুটির মধ্যে অন্ততঃ একটিও ঘটে। কোন পরীক্ষণ $\varepsilon$-এর সংশ্লিষ্ট মৌলিক ঘটনাবলী $w_i$ $(i = 1, 2, 3, \ldots)$ এর সমবায়ে দুটি ঘটনা $A = \{w_1, w_2, w_3\}$ ও $B = \{w_1, w_3, w_4, w_5, w_6\}$ গঠিত হলে $A \cup B = \{w_1, w_2, w_3, w_4, w_5, w_6\}$ নির্দেশ করবে সেই ঘটনা যা ঘটবে যদি $w_1, \ldots, w_6$ এই মৌলিক ঘটনা-কয়টির যে কোন একটি ঘটে। ছক্কা নিক্ষেপণের পরীক্ষণে $A = \{3, 6\}$ ও $B = \{2, 4, 6\}$ যথাক্রমে 3-এর গুণনীয়ক এবং যুগ্মসংখ্যা নির্দেশক চিহ্নের আবির্ভাব বোঝায়, তবে $A \cup B = \{2, 3, 4, 6\}$ বোঝাবে 3-এর গুণনীয়ক এবং/অথবা যুগ্মরাশি-নির্দেশক চিহ্নের আবির্ভাব। সাধারণভাবে, যদি $A_1, \ldots, A_i, \ldots, A_n$ যে কোন পরীক্ষণ-সংশ্লিষ্ট $n$ টি ঘটনা হয়, তবে $\bigcup_{i=1}^{n} A_i$ সংকেত-চিহ্ন ব্যবহার ক'রে আমরা নির্দেশ করব সেই ঘটনা যা ঘটবে যদি এদের মধ্যে অন্ততঃ একটি ঘটনাও ঘটে। $A$ ও $B$ দুটি পরস্পরব্যতিরেকী ঘটনা হলে

$A \cup B$-এর বিকল্পে $A+B$ এবং $A_1, \ldots, A_i, \ldots A_n$ যৌথভাবে পরস্পরব্যতিরেকী ঘটনা হলে $\bigcup_{i=1}^{n}$-এর বিকল্পে $\sum_{i=1}^{n} A_i = A_1 + \cdots + A_n$ সংকেতসূত্রও ব্যবহার করব।

$A$ ও $B$ যদি এমন দুটি ঘটনা হয় যে, $A$ ঘটনাটি ঘটলে $B$ ঘটনাটি ঘটবেই কিন্তু $B$ ঘটনাটি ঘটলে $A$ ঘটতেও পারে বা নাও ঘটতে পারে, তাহলে আমরা লিখব $A \subset B$ এবং বলব যে $A$ ঘটনার সংঘটন $B$ ঘটনার আবশ্যিক সংঘটন সূচিত করে। একটি ছক্কায় 6-এর গুণনীয়ক সংখ্যানির্দেশক চিহ্ন দৃষ্ট হওয়াকে $A$ এবং যুগ্মরাশি নির্দেশক চিহ্ন দৃষ্ট হওয়াকে যদি $B$ বলি, তবে $A=\{6\}$ ও $B=\{2, 4, 6\}$ এবং স্পষ্টতঃই $A \subset B$. যদি $A \subset B$ সত্যি না হয়, তবে লিখব $A \not\subset B$. ওপরের উদাহরণে $A \subset B$; কিন্তু $B \not\subset A$. যদি $A \subset B$ এবং $B \subset A$ উভয়েই একযোগে সত্য হয়, অর্থাৎ যদি এমন হয় যে $A$ ঘটলে $B$ এবং $B$ ঘটলে $A$ ঘটবেই এমন পরিস্থিতি দাঁড়ায়, তাহলে এই ঘটনা-দুটিকে সমতুল্য (equivalent) বলা হয় এবং তখন আমরা $A=B$ লিখব। যদি ছক্কার ওপর 3-এর অখণ্ড গুণনীয়ক সংখ্যার আবির্ভাবকে $A$ দ্বারা নির্দেশ করা হয় অর্থাৎ যদি $A=\{3, 6\}$ হয়, তবে $A \not\subset B$ এবং $B \not\subset A$. আবার, $A-B$ সংকেতচিহ্ন ব্যবহার ক'রে বোঝানো হয় সেই ঘটনা যা ঘটবে যদি, এবং কেবলমাত্র যদি, $A$ ঘটনাটি ঘটে কিন্তু $B$ ঘটনাটি না ঘটে। মনে কর, ছক্কা নিক্ষেপণে 3-এর গুণনীয়ক ও যুগ্ম রাশিসূচক চিহ্নের আবির্ভাব-ঘটনা হচ্ছে যথাক্রমে $A$ এবং $B$ অর্থাৎ $A=\{3, 6\}$ এবং $B=\{2, 4, 6\}$. তাহলে, $A-B=\{3\}$. অর্থাৎ এটি নির্দেশ করে 3-এর গুণনীয়ক অযুগ্ম সংখ্যানির্দেশক চিহ্নের আবির্ভাব।

যে কোন $n$ সংখ্যক পৃথক্ ঘটনা $A_1, A_2, \ldots, A_i, \ldots, A_n$-কে একত্রযোগে 'পরস্পরব্যতিরেকী ও পরস্পরনিঃশেষী ঘটনাপুঞ্জ' (mutually exclusive and mutually exhaustive set of events) বলা হয় যদি তারা যৌথভাবে পরস্পর ব্যতিরেকী হয় এবং সঙ্গে সঙ্গে তাদের মধ্যে অন্ততঃ একটি ঘটনা যে ঘটবেই তা যেন নিশ্চিত ঘটনা হয়। অর্থাৎ এই $n$ সংখ্যক ঘটনাবলীর বৈশিষ্ট্য হ'ল দুটি :

$$\text{প্রত্যেক } i, j = 1, 2, \ldots n \ (i \neq j)\text{-এর জন্যে } A_i \cap A_j = \phi \qquad (7.4)$$

$$\text{এবং } \bigcup_{i=1}^{n} A_i = \sum_{i=1}^{n} A_i = \Omega. \qquad \ldots \qquad \ldots \qquad (7.5)$$

ফলে, প্রত্যেক $i, j = 1, \ldots, n$ $(i \neq j)$-এর জন্য $P(A_i \cap A_j) = 0$ (7.6)

এবং $P(\bigcup_{i=1}^{n} A_i) = P(\sum_{i=1}^{n} A_i) = P(\Omega) = 1.$ (7.7)

যদি দুটি ঘটনা $A$ এবং $B$ পরস্পর এমন সম্বন্ধযুক্ত হয় যে, এদের মধ্যে একটি ঘটলে আর একটি ঘটতে পারে না এবং একটি না ঘটলে অপরটি ঘটতে বাধ্য, তবে এদের একটিকে অপরটির পরিপূরক (complementary) ঘটনা বলা হয়। কোন পরীক্ষণ ε-এর সঙ্গে সংশ্লিষ্ট যে মৌলিক ঘটনাবলীর সমবায়ে $A$ ঘটনাটি গঠিত হবে এক্ষেত্রে পরিপূরক $B$ ঘটনাটি গঠিত হবে সেগুলি ছাড়া বাকী সমস্ত মৌলিক ঘটনাপুঞ্জের সমন্বয়ে। $B$ ঘটনাটি $A$ ঘটনার পরিপূরক হলে সাধারণতঃ আমরা লিখব $B = A^*$. তাহলে, $A$ এবং তার পরিপূরক $A^*$ ঘটনার পারস্পরিক সম্পর্ক দাঁড়ালো এই যে,

(1) $A$ এবং $A^*$-এর একত্র সংঘটন একটি অসম্ভব ঘটনা অর্থাৎ $A \cap A^* = \phi$

অর্থাৎ $P(A \cap A^*) = 0$ ... ... (7.8)

এবং (2) $A$ ও $A^*$-এর মধ্যে একটি যে ঘটবেই তা হচ্ছে একটি নিশ্চিত ঘটনা অর্থাৎ $A \cup A^* = A + A^* = \Omega$ অর্থাৎ

$P(A \cup A^*) = 1.$ ... ... (7.9)

সংক্ষেপে, $A$ এবং $A^*$ একত্রে পরস্পরব্যতিরেকী ও পরস্পরনিঃশেষী।

## 7.5 কয়েকটি উপপাদ্য ও অনুসিদ্ধান্ত :

**উপপাদ্য 1.** সামগ্রিক সম্ভাবনা উপপাদ্য বা সম্ভাবনার যৌগিক উপপাদ্য (theorem of total probability or addition theorem of probability).

**নির্বচন :** যে কোন পরস্পরব্যতিরেকী $k$ সংখ্যক ঘটনা $A_1, \ldots, A_i, \ldots,$ $A_k$-এর মধ্যে অন্ততঃ একটির সংঘটন সম্ভাবনা হচ্ছে এই ঘটনাগুলির পৃথক্ পৃথক্ সম্ভাবনা সমূহের সমষ্টি।

সংকেতচিহ্ন ব্যবহার ক'রে বলা যায়

$$P\left(\sum_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} P(A_i). \quad \cdots \quad \cdots \quad (7.10)$$

**প্রমাণ :** মনে কর $A_1, A_2, \cdots, A_i, \ldots, A_k$ ঘটনাগুলি একটি পরীক্ষণ ε-এর সঙ্গে সংশ্লিষ্ট এবং পরীক্ষণের মোট সমসম্ভব পরিস্থিতিসংখ্যা $N$. ধর $A_1, \ldots, A_i,$

$\ldots, A_k$ ঘটনাগুলির পৃথক্ পৃথক্ অনুকূল পরিস্থিতিসংখ্যা যথাক্রমে $n_1, \ldots, n_i, \ldots, n_k$. তাহলে, ঘটনাগুলি সব পরস্পরব্যতিরেকী হওয়ার ফলে '$A_1$ অথবা $A_2$ অথবা $A_3, \ldots,$ অথবা $A_k$' এই ঘটনাটির, অর্থাৎ $\sum_{i=1}^{k} A_i$ ঘটনাটির, অনুকূলে মোট পরিস্থিতিসংখ্যা হচ্ছে $\sum_{i=1}^{n} n_i$, কারণ $n_i$ সংখ্যক বিভিন্ন পরিস্থিতি যা $A_i$-এর অনুকূলে রয়েছে তা অন্য কোন ঘটনা $A_j$ এর অনুকূলে নেই ( যদি $j \neq i = 1, \ldots, k$ হয় )। কাজেই সম্ভাবনার পুরাতনী সংজ্ঞানুসারে,

$$P\left(\sum_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} n_i/N = \sum_{i=1}^{k}\left(\frac{n_i}{N}\right) = \sum_{i=1}^{k} P(A_i). \qquad (7.11)$$

**অনুসিদ্ধান্ত 1.** যদি $k$-সংখ্যক পরস্পরব্যতিরেকী ঘটনা $A_1, \ldots, A_i, \ldots, A_k$-এর মধ্যে যে কোন একটি ঘটলে বলা হয় যে $A$ ঘটনাটি ঘটেছে অর্থাৎ যদি $A$ ঘটনাটি $k$-সংখ্যক বিভিন্ন রূপে (in different forms) ঘটে, তবে $A$ ঘটনার সম্ভাবনা হচ্ছে $A_1, \ldots, A_i, \ldots, A_k$ ঘটনাগুলির পৃথক্ পৃথক্ সম্ভাবনার সমষ্টি।

**প্রমাণ :** সংজ্ঞানুযায়ী, $A = \sum_{i=1}^{k} A_i$. কাজেই $P(A) = P\left(\sum_{i=1}^{k} A_i\right)$,

আবার, যেহেতু, $P\left(\sum_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} P(A_i)$ ( 7.11 দ্রষ্টব্য )

$$\text{সুতরাং,} \quad P(A) = \sum P(A_i). \qquad \cdots \qquad \cdots \qquad (7.12)$$

2. যদি $A$ ও $A^*$ পরস্পরের পরিপূরক ঘটনা হয়,

$$\text{তবে} \quad P(A^*) = 1 - P(A). \qquad \cdots \qquad \cdots \qquad (7.13)$$

**প্রমাণ :** দেওয়া আছে $A \cap A^* = \phi$ এবং $A \cup A^* = A + A^* = \Omega$.

সুতরাং $1 = P(\Omega) = P(A) + P(A^*)$ [ (7.11) দ্রষ্টব্য ]

অর্থাৎ $P(A^*) = 1 - P(A)$.

**উপপাদ্য 2.** $A$ ও $B$ যদি যে কোন দুটি ঘটনা হয়, তবে

$$P(A \cup B) = P(A) + P(B) - P(A \cap B). \qquad \cdots \qquad (7.14)$$

**প্রমাণ :** $A$ ঘটনাটি $A \cap B$ এবং $A \cap B^*$—এই দুটি পরস্পরব্যতিরেকী ঘটনার একটি ঘটলে তবেই ঘটবে, কারণ $A$ ঘটতে পারে হয় অপর একটি ঘটনা $B$-এর সঙ্গে একত্রে অথবা $B^*$-এর সঙ্গে একত্রে। অর্থাৎ $A$ দুটি পরস্পর-ব্যতিরেকী রূপে, যথা (1) $A \cap B$ রূপে অথবা (2) $A \cap B^*$ রূপে ঘটতে পারে। তাই আমরা লিখতে পারি

$$A = [A \cap B] + [A \cap B^*] \quad \cdots \quad \cdots \quad (7.15)$$

সুতরাং $P(A) = P[A \cap B] + [A \cap B^*] = P[A \cap B] + P[A \cap B^*]$

$$\cdots \quad \cdots \quad (7.16)$$

আবার, (7.15)-এর মত আমরা লিখতে পারব

$$B = [A \cap B] + [A^* \cap B] \quad \cdots \quad \cdots \quad (7.17)$$

সুতরাং $P(B) = P(A \cap B) + P(A^* \cap B) \quad \cdots \quad \cdots \quad (7.18)$

তাহলে, (7.16) ও (7.18) থেকে পাই

$$P(A) + P(B) = P(A \cap B) + [P(A \cap B^*) + P(A \cap B) + P(A^* \cap B)] \quad (7.19)$$

এটা সহজেই বোধগম্য যে, $A \cap B^*$, $A \cap B$ এবং $A^* \cap B$ হচ্ছে পরস্পর-ব্যতিরেকী ঘটনা।

সুতরাং, $P([A \cap B^*] + [A \cap B] + [A^* \cap B])$

$$= P[A \cap B^*] + P[A \cap B] + P[A^* \cap B]. \quad \cdots \quad (7.20)$$

এখন $A \cap B^* + A \cap B + A^* \cap B$ হচ্ছে সেই ঘটনা যা ঘটবে তিনটি পরস্পরব্যতিরেকী রূপে, যথা : (1) $A \cap B$ রূপে, (2) $A \cap B^*$ রূপে এবং (3) $A^* \cap B$ রূপে। আবার, $A \cup B$ ঘটনাটিও ঠিক এই তিনটি বিভিন্ন পরস্পরব্যতিরেকী রূপেই ঘটতে পারে।

কাজেই $A \cup B = [A \cap B^*] + [A \cap B] + [A^* \cap B]$.

সুতরাং $P[A \cap B^* + A \cap B + A^* \cap B] = P(A \cup B). \quad \cdots \quad (7.21)$

কাজেই (7.19) – (7.21) থেকে পাই

$$P(A) + P(B) = P(A \cap B) + P(A \cup B)$$

অর্থাৎ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

**উপপাদ্য 3.** $A_1, \ldots, A_i, \ldots A_m$ যদি $m$-সংখ্যক বিভিন্ন ঘটনা হয়, তবে

$$P\left(\bigcup_{i=1}^{m} A_i\right) = \sum_{i=1}^{m} P(A_i) - \sum_{i<j=1}^{m} P(A_i \cap A_j)$$

$$+ \sum_{i<j<k=1}^{m} P(A_i \cap A_j \cap A_k) -$$

$$\cdots + (-1)^{m-1} P(A_1 \cap A_2 \cap \ldots \cap A_m). \quad (7.22)$$

**প্রমাণ :** উপপাদ্য 2 থেকে পাই

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

অর্থাৎ $P\left(\bigcup_{i=1}^{2} A_i\right) = \sum_{i=1}^{2} P(A_i) + (-1) P(A_1 \cap A_2).$

কাজেই, $m = 2$-এর বেলায় উপপাদ্য 3 সত্য।

আবার, $P\left(\bigcup_{i=1}^{3} A_i\right) = P(A_1 \cup A_2 \cup A_3) = P([A_1 \cup A_2] \cup A_3)$

$= P(A_1 \cup A_2) + P(A_3) - P([A_1 \cup A_2] \cap A_3)$ [ উপপাদ্য 2 দ্রষ্টব্য ]

$= P(A_1 \cup A_2) + P(A_3) - P([A_1 \cap A_3] \cup [A_2 \cap A_3])$

[ কারণ, $[A_1 \cup A_2] \cap A_3 = [A_1 \cap A_3] \cup [A_2 \cap A_3]$, যেহেতু বামপক্ষ নির্দেশ করছে $A_1$ অথবা $A_2$ বা তাদের উভয়ের সঙ্গে $A_3$ ঘটনার একত্র সংঘটন এবং দক্ষিণপক্ষ নির্দেশ করছে $A_1$ ও $A_3$ এবং/অথবা $A_2$ ও $A_3$-এর একত্র সংঘটন ; কাজেই উভয়পক্ষই একই ঘটনা নির্দেশ করছে। ]

$= P(A_1 \cup A_2) + P(A_3) - [P(A_1 \cap A_3) + P(A_2 \cap A_3)$

$- P([A_1 \cap A_3] \cap [A_2 \cap A_3])]$ [ উপপাদ্য 2 দ্রষ্টব্য ]

$= P(A_1) + P(A_2) - P(A_1 \cap A_2) + P(A_3) - P(A_1 \cap A_3)$

$- P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$

[ উপপাদ্য 2 দ্রষ্টব্য এবং লক্ষণীয় যে,

$(A_1 \cap A_3) \cap (A_2 \cap A_3) = A_1 \cap A_2 \cap A_3$ ]

$$= \sum_{i=1}^{3} P(A_i) - \sum_{i<j=1}^{3} P(A_i \cap A_j) + (-1)^{3-1} P(A_1 \cap A_2 \cap A_3).$$

কাজেই $m=3$-এর বেলায়ও উপপাদ্য 3 সত্য।

এখন ধ'রে নেওয়া যাক যে উপপাদ্য 3 যে কোন অখণ্ড ধনাত্মক সংখ্যা $m$-এর জন্যে সত্য। তাহলে $(m+1)$-এর জন্যে পাই

$$P(\bigcup_{i=1}^{m+1} A_i)=P([\bigcup_{i=1}^{m} A_i]\cup A_{m+1})$$

$$=P(\bigcup_{i=1}^{m} A_i)+P(A_{m+1})-P([\bigcup_{i=1}^{m} A_i]\cap A_{m+1})$$

$$=P(\bigcup_{i=1}^{m} A_i)+P(A_{m+1})-P(\bigcup_{i=1}^{m}[A_i\cap A_{m+1}])$$

$$=\sum_{i=1}^{m} P(A_i)-\sum_{i<j=1}^{m} P(A_i\cap A_j)$$

$$+\sum_{i<j<k=1}^{m} P(A_i\cap A_j\cap A_k)-\cdots$$

$$+(-1)^{m-1}P(A_1\cap A_2\ldots\cap A_m)]+P(A_{m+1})$$

$$-\Big[\sum_{i=1}^{m} P(B_i)-\sum_{i<j=1}^{m} P(B_i\cap B_j)$$

$$+\sum_{i<j<k=1}^{m} P(B_i\cap B_j\cap B_k)-\cdots\cdots$$

$$+(-1)^{m-1}P(B_1\cap B_2\cap\ldots\cap B_m)\Big]$$

[ $B_i=A_i\cap A_{m+1}$, $i=1, \ldots, m$ লিখে এবং (7.22) থেকে ]

$$=\Big[\sum_{i=1}^{m} P(A_i)-\sum_{i<j=1}^{m} P(A_i\cap A_j)$$

$$+\sum_{i<j<k=1}^{m} P(A_i\cap A_j\cap A_k)-\cdots$$

$$+(-1)^{m-1}P(A_1\cap A_2\cap\ldots\cap A_m)]+P(A_{m+1})$$

$$-\Big[\sum_{i=1}^{m} P(A_i \cap A_{m+1}) - \sum_{i<j=1}^{m} P(A_i \cap A_j \cap A_{m+1})$$

$$+ \sum_{i<j<k=1}^{m} P(A_i \cap A_j \cap A_k \cap A_{m+1}) - \cdots$$

$$+ (-1)^{m-1} P(A_1 \cap A_2 \cap \ldots \cap A_{m+1})\Big]$$

অর্থাৎ উপপাদ্য 3, $m$-এর জন্যে সত্য হলে $(m+1)$-এর জন্যেও সত্য হবে। কিন্তু আগে দেখেছি যে এটি $m=2$ এবং $m=3$-এর জন্যে খাটে। কাজেই এটি $m=4, 5, 6, \cdots$ইত্যাদি সকল অখণ্ড ধনরাশির জন্যেই খাটে। কাজেই আরোহ পদ্ধতি (method of induction) অনুসরণ ক'রে উপপাদ্যটি এভাবে প্রমাণিত হ'ল।

**অনুসিদ্ধান্ত**: $P(\underset{i=1}{\cup} A_i) \leqslant \underset{i=1}{\Sigma} P(A_i)$ (7.23)

**প্রমাণ ঃ** $P(A_1 \cup A_2) = P(A_1) + P(A_2)$

$$- P(A_1 \cap A_2) \leqslant P(A_1) + P(A_2)$$

$$P(A_1 \cup A_2 \cup A_3) = P(A_1 \cup A_2) + P(A_3) - P([A_1 \cup A_2] \cap A_3)$$

$$\leqslant P(A_1 \cup A_2) + P(A_3) \leqslant P(A_1) + P(A_2) + P(A_3)$$ ইত্যাদি।

[ এর পর আরোহ পদ্ধতি প্রয়োগ ক'রে প্রমাণটি নিজে শেষ কর। ]

## 7.6 কয়েকটি উদাহরণ ঃ

**উদা 7.8** একটি দাবাখেলার ছকে 64টি বর্গাকৃতি খোপ থেকে সমসম্ভব উপায়ে 3টি বেছে নিলে তারা কোণাকুণি বিন্যস্ত হবে এরূপ সম্ভাবনা কত ?

এখানে আলোচ্য ঘটনাটি 22টি বিভিন্ন পরস্পরব্যতিরেকী রূপে ঘটতে পারে। ছকটিতে 2টি কোণাকুণি বিন্যস্ত 8 খোপের গুচ্ছ এবং 4টি ক'রে কোণাকুণি বিন্যস্ত 3, 4, 5, 6 এবং 7 খোপের গুচ্ছ রয়েছে। এই 22টি গুচ্ছের যে কোন একটি থেকে যদি 3টি খোপ বেছে নেওয়া হয় তাহলেই প্রশ্ননির্দিষ্ট ঘটনাটি ঘটবে। এখানে মোট সমসম্ভব পরিস্থিতিসংখ্যা স্পষ্টতঃই $\binom{64}{3}$ এবং ঘটনাটির অনুকূল পরিস্থিতিসংখ্যা হচ্ছে

$$2 \times \binom{8}{3} + 4\Big[\binom{3}{3} + \binom{4}{3} + \binom{5}{3} + \binom{6}{3} + \binom{7}{3}\Big] = 392.$$

সুতরাং নির্ণেয় সম্ভাবনা $= \dfrac{392}{\binom{64}{3}} = \dfrac{7}{744}.$

**উদা 7.9** দুটি অখণ্ড ধনরাশি যদি সমসম্ভব উপায়ে বেছে নেওয়া হয় যাতে তাদের সমষ্টি 100-এর সমান থাকে, তাহলে তাদের গুণফল 1000-এর চেয়ে বেশী হবার সম্ভাবনা নির্ণয় কর।

ধরা যাক, একটি সংখ্যা $x$; তবে অপরটি $100-x$. এখানে পরীক্ষণ হচ্ছে সমসম্ভব উপায়ে 1 থেকে 99-এর মধ্যবর্তী একটি অখণ্ড ধনরাশিকে $x$-এর মান হিসেবে বেছে নেওয়া। কাজেই মোট পরিস্থিতিসংখ্যা 99. প্রদত্ত ঘটনাটি ঘটতে হলে $x$-এর মান এমন হওয়া চাই যেন $x(100-x)>1000$ হয় অর্থাৎ যেন $(x-50)^2<1500$ হয়। তাহলে $x$-এর মান $12, 13, \ldots, 88$ হলে তবেই এই সর্তটি খাটবে। কাজেই ঘটনার অনুকূল পরিস্থিতিসংখ্যা 77 এবং নির্ণেয় সম্ভাবনা $\frac{77}{99}=\frac{7}{9}$.

**উদা 7.10** প্রথম $n$-সংখ্যক অখণ্ড ধনরাশি $1, 2, \ldots, i, \ldots, n$-কে যদি সমসম্ভব উপায়ে পরপর সাজানো যায় এবং যে স্থানগুলিতে তারা বসবে সেগুলিকে যদি $1, 2, \ldots, i, \ldots, n$ সংখ্যা দিয়ে চিহ্নিত করা হয়, তাহলে কোন রাশিই অনুরূপ সংখ্যা-চিহ্নিত স্থানে না বসবার সম্ভাবনা কত?

$i$-সংখ্যাটি $i$-চিহ্নিত স্থানে বসলে আমরা বলব যে $E_i$ ঘটনাটি ঘটেছে। তাহলে $E=\bigcup_{i=1}^{n} E_i$ হচ্ছে সেই ঘটনা যা ঘটবে যদি অন্ততঃ একটি রাশিও স্ব-সংখ্যক স্থানে বসে। তাহলে $E$-এর পরিপূরক ঘটনা $E^*$ হচ্ছে কোন রাশিই তদনুগ সংখ্যা চিহ্নিত স্থানে না বসার ঘটনা। এবং আমরা $E^*$-এর সম্ভাবনা নির্ণয় করতে চাই।

$$\text{এখন,}\quad P(E^*)=1-P(E)=1-P\left(\bigcup_{i=1}^{n} E_i\right)$$

$$=1-\sum_{i=1}^{n} P(E_i)+\sum_{i<j=1}^{n} P(E_i\cap E_j)-\cdots$$

$$+(-1)^{n-1}\times(-1)P(E_1\cap E_2\cap\ldots\cap E_n)$$

এখন, $P(E_i)=\dfrac{(n-1)!}{n!}$ কারণ $n$ রাশিকে $n$ সংখ্যক স্থানে অবাধে $n!$ উপায়ে বসানো যায় এবং $i$-সংখ্যাটি $i$-চিহ্নিত স্থানে বসলে বাকী $(n-1)$ রাশিকে বাকি $(n-1)$ স্থানে অবাধে $(n-1)!$ উপায়ে বসানো যায়। একই যুক্তিতে

$$P(E_i \cap E_j) = \frac{(n-2)!}{n!},\ P(E_i \cap E_j \cap E_k) = \frac{(n-3)!}{n!},\ \ldots \text{ইত্যাদি।}$$

এবং সর্বশেষে $P(E_1 \cap E_2 \cap \ldots \cap E_n) = \frac{1}{n!}$.

সুতরাং
$$P(E^*) = 1 - n \times \frac{1}{n} + \binom{n}{2} \times \frac{1}{n(n-1)}$$
$$- \binom{n}{3} \times \frac{1}{n(n-1)(n-2)} + \cdots + (-1)^n \binom{n}{n} \frac{1}{n!}$$
$$= 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots + (-1)^n \cdot \frac{1}{n!}$$
$$= \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots + (-1)^n \cdot \frac{1}{n!}.$$

## 7.7 সর্তাধীন সম্ভাবনা ও ঘটনার স্বাতন্ত্র্য (conditional probability and independence of events) :

মনে কর, একটি ঘটনা $A$-র সম্ভাবনা $P(A)$-র মান ধনাত্মক এবং $B$ অপর একটি ঘটনা। এখন যদি কোন উপায়ে জানা যায় যে $A$ ঘটনাটি পূর্বেই ঘটে গিয়েছে তাহলে এই তথ্য সম্বন্ধে উদাসীন থেকে $B$-এর সম্ভাবনা $P(B)$-এর মান যা হবে তা $A$ ঘটনা পূর্বে ঘটে গেছে, এই অতিরিক্ত তথ্য ব্যবহার ক'রে $B$ ঘটনার সম্ভাবনার যে মান পাওয়া যেতে পারে তার সঙ্গে সমান নাও হতে পারে। ধনাত্মক সম্ভাবনাযুক্ত কোন ঘটনা $A$ পূর্বে ঘটে গেছে এই সর্তসাপেক্ষে $B$ ঘটনার সম্ভাবনাকে $B$ ঘটনার সর্তাধীন সম্ভাবনা (conditional probability) বলে। এখানে সর্ত হচ্ছে এই যে, $P(A) > 0$ এবং $A$ ঘটনা পূর্বে ঘটে গেছে এবং এই অতিরিক্ত তথ্য $B$-এর সম্ভাবনা নির্ণয়ে ব্যবহৃত হয়েছে। এই সর্তাধীন সম্ভাবনাকে $P(B|A)$ বা $P_A(B)$ সংকেত সূত্রে প্রকাশ করা হবে। সাধারণতঃ $P(B|A)$-এর মান $P(B)$ থেকে পৃথক্, যদিও সব সময় নয়। এখানে $B|A$ সংকেতচিহ্ন ব্যবহার ক'রে $A$ ঘটনার প্রাক্‌সংঘটন সাপেক্ষে $B$-এর সর্তাধীন ঘটনা (conditional event) বোঝানো হয়।

**উপপাদ্য 4.** মিশ্রসম্ভাবনা উপপাদ্য (theorem of compound probability).

**নির্বচন :** দুটি ঘটনা $A$ এবং $B$-এর জন্যে যদি দেওয়া থাকে যে

$P(A) > 0$ এবং $P(B|A)$ হচ্ছে $A$ পূর্বে ঘটে গেছে এই সর্তাধীনে $B$-এর সম্ভাবনা, তাহলে

$$P(A \cap B) = P(A).P(B|A). \quad \cdots \quad (7.24)$$

**প্রমাণ :** মনে কর, $A$ ও $B$ ঘটনা-দুটি একটি পরীক্ষণ $\varepsilon$-এর সঙ্গে সংশ্লিষ্ট। ধর, পরীক্ষণটিতে মোট সমসম্ভব পরিস্থিতিসংখ্যা $N$ এবং তার মধ্যে $A$ এবং $A \cap B$ ঘটনা-দুটির অনুকূলে আছে যথাক্রমে $N(A)$ এবং $N(A \cap B)$ সংখ্যক পরিস্থিতি। স্পষ্টতঃই $N(A \cap B) \leqslant N(A) \leqslant N$. তাহলে সংজ্ঞানুযায়ী

$$P(A \cap B) = \frac{N(A \cap B)}{N} = \frac{N(A)}{N} \cdot \frac{N(A \cap B)}{N(A)} \quad \cdots \quad (7.25)$$

আমরা এভাবে লিখতে পারি যেহেতু $N(A) > 0$, কারণ $\frac{N(A)}{N} = P(A) > 0$.

আবার স্পষ্টতঃই $\frac{N(A \cap B)}{N(A)}$ হচ্ছে $P(B|A)$-এর সমান। কারণ, যদি এটা স্বীকার করা হয় যে, $A$ ঘটনাটি ঘটে গেছে, তাহলে মোট $N$টি মৌলিক ঘটনার মধ্যে এখন কেবল $N(A)$ সংখ্যক মৌলিক ঘটনাই সম্ভব (likely) বলে স্বীকার্য। আবার এই মৌলিক ঘটনাগুলি সমসম্ভবও বটে এবং এদের মধ্যে $N(A \cap B)$টি পরিস্থিতি হচ্ছে $B$ ঘটনারও অনুকূল। কাজেই আমরা লিখতে পারি

$$\frac{N(A)}{N} \cdot \frac{N(A \cap B)}{N(A)} = P(A).P(B|A) \quad \cdots \quad (7.26)$$

সুতরাং (7.25) ও (7.26) থেকে পাই

$$P(A \cap B) = P(A).P(B|A).$$

**অনুসিদ্ধান্ত :** যদি $A$, $B$ ও $C$ তিনটি ঘটনা হয় এবং $P(A \cap B) > 0$ হয়, তাহলে

$$P(A \cap B \cap C) = P(A).P(B \cap C|A) = P(A).P(B|A).P(C|A \cap B). \quad (7.27)$$

এখানে $C|A \cap B$ হচ্ছে $A$ ও $B$-এর যুগপৎ প্রাক্‌সংঘটন সাপেক্ষে $C$-এর সর্তাধীন সংঘটন।

**টীকা :** মিশ্রসম্ভাবনা উপপাদ্য থেকে $A$ ঘটনার প্রাক্‌সংঘটন সাপেক্ষে $B$-এর সর্তাধীন সম্ভাবনাকে লেখা যায়

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \; P(A) > 0 \text{ হলে।} \quad \cdots \quad (7.28)$$

এখানে $P(A) > 0$ হবেই কারণ $P(A) \geqslant P(A \cap B) > 0$.

অনুরূপভাবে, $B$ ঘটনার প্রাক্‌সংঘটন সাপেক্ষে $A$-এর সর্তাধীন সম্ভাবনা হচ্ছে

$$P(A|B)=\frac{P(A\cap B)}{P(B)},\ P(B)>0 \text{ হলে।} \quad \cdots \quad (7.29)$$

এখানেও $P(B)>0$ হবেই কারণ $P(B)\geqslant P(A\cap B)>0$.

যদি $A$ ঘটনার প্রাক্‌সংঘটন সাপেক্ষে $B$ ঘটনার সর্তাধীন সম্ভাবনা $P(B|A)$, $B$ ঘটনার নিঃসর্ত সম্ভাবনা অর্থাৎ $P(B)$-এর সমান হয়, তাহলে $B$-কে $A$ থেকে স্বতন্ত্র বা $A$-র অনধীন (independent of $A$) বলা হয়। এই স্বাতন্ত্র্য বা অনধীনতা হচ্ছে সম্ভাবনাতাত্ত্বিক (stochastic বা probabilistic) অর্থে। সংক্ষেপে, $B$ সম্ভাবনাতত্ত্বগত অর্থে $A$-এর অনধীন হবে যদি

$$P(B)=P(B|A)=\frac{P(A\cap B)}{P(A)} \text{ হয়} \quad (7.30)$$

অর্থাৎ যদি $P(A\cap B)=P(A).P(B)$ হয়। ... ... (7.31)

তেমনিভাবে সম্ভাবনাতাত্ত্বিক অর্থে $A$ ঘটনা $B$ ঘটনার অনধীন হবে, যদি

$$P(A)=P(A|B)=\frac{P(A\cap B)}{P(B)} \text{ হয়, ( } P(B)>0 \text{ ধ'রে )} \quad (7.32)$$

অর্থাৎ যদি $P(A\cap B)=P(A).P(B)$ হয় (7.33)

এখন (7.31) ও (7.33)-এর অভিন্নতা লক্ষ্য ক'রে বলা যায় যে, সম্ভাবনা-তত্ত্বানুযায়ী $A$ ও $B$ ঘটনাদ্বয় পরস্পর স্বতন্ত্র বা অনধীন (stochastically mutually independent) হবে যদি $P(A\cap B)=P(A).P(B)$ হয়। একেই $A$ ও $B$ ঘটনাদ্বয়ের সম্ভাবনাতাত্ত্বিক অনধীনতার সংজ্ঞা হিসেবে নেওয়া হবে। এখানে উল্লেখ্য যে, (7.31) ও (7.33) যথাক্রমে (7.30) ও (7.32) থেকে অনুসৃত। কিন্তু (7.30) ও (7.32) এর সত্যতা যথাক্রমে $P(A)>0$ ও $P(B)>0$ এর সত্যতার ওপর নির্ভর করছে। কিন্তু (7.31) বা (7.33)-কে $A$ ও $B$-এর অনধীনতার সংজ্ঞা হিসেবে নেওয়া চলে যদি $P(A)>0$ বা $P(B)>0$ সত্য নাও হয়। $P(A)$ বা $P(B)$-এর মান 0 হলেও (7.31) বা (7.33) সম্পর্কটি সত্য; কারণ $P(A)$ বা $P(B)$ শূন্য হলে $P(A\cap B)$-এর মান শূন্য হবেই যেহেতু $P(A)$ [ বা $P(B)$ ] শূন্য হওয়ার অর্থ এই যে $A$ [ বা $B$ ] ঘটনার অনুকূল পরিস্থিতিসংখ্যা শূন্য এবং সেক্ষেত্রে $A\cap B$-এর অনুকূল পরিস্থিতিসংখ্যাও শূন্য হতে বাধ্য। যদি (7.31) বা (7.33) সত্য না হয় তাহলে বলা হবে যে ঘটনা-দুটি সম্ভাবনাসূত্রে পরস্পর নির্ভরশীল (stochastically interdependent). বাস্তবিক, একটি ঘটনা আগে ঘটে গেছে, এই তথ্য ব্যবহার

ক'রে অপর একটি ঘটনার সর্তাধীন সম্ভাবনা এবং এই তথ্য সম্পর্কে উদাসীন থেকে শেষোক্ত ঘটনাটির সর্তনিরপেক্ষ সম্ভাবনা যদি ভিন্ন মানসম্পন্ন হয়, তবে এটা বলা স্বাভাবিক যে ঐ ঘটনাটি প্রথমোক্ত ঘটনার সংঘটনের ওপর নির্ভরশীল। এই নির্ভরশীলতা হচ্ছে সম্ভাবনাতাত্ত্বিক নির্ভরশীলতা (stochastic dependence). যেমন, $P(B|A)$ ও $P(B)$-এর মান পৃথক্ হলে বলা হবে যে $B$ সম্ভাবনাসূত্রে $A$-এর ওপর নির্ভরশীল। এক্ষেত্রে অবশ্যই $P(A|B)$ ও $P(A)$-এর মানও পৃথক্ হতে বাধ্য [ (7.30) ও (7.32) দ্রষ্টব্য ] এবং ফলে $A$ সম্ভাবনাগতভাবে $B$-এর অধীন। বাস্তবিক, $A$ ও $B$ উভয়েই পরস্পর নির্ভরশীল।

তিনটি পৃথক্ ঘটনা $A$, $B$ ও $C$ সম্ভাবনাগতভাবে পরস্পর নির্ভরতাশূন্য হবে যদি

$$\left.\begin{aligned} P(A\cap B\cap C) &= P(A)\,P(B)\,P(C),\\ P(A\cap B) &= P(A)\,P(B)\\ P(A\cap C) &= P(A)\,P(C)\\ \text{এবং }\; P(B\cap C) &= P(B)\,P(C) \end{aligned}\right\} \quad \cdots \quad (7.34)$$

সত্য হয়।

যে কোন $n$ সংখ্যক পৃথক্ ঘটনা $A_1, \ldots, A_i, \ldots, A_n$-এর জন্যে যদি $P(A_1\cap \ldots \cap A_{n-1}) > 0$ হয় এবং $P(A_k|A_1\cap A_2 \ldots \cap A_{k-1})$, $k=2, 3, \ldots n$, যদি $A_1, \ldots, A_{k-1}$-এ যুগপৎ প্রাক্‌সংঘটন সাপেক্ষে $A_k$-এর সর্তাধীন সম্ভাবনা নির্দেশ করে, তবে দেখানো যায় যে,

$$\begin{aligned} &P(A_1\cap \ldots \cap A_n)\\ &= P(A_1)\,P(A_2|A_1)P(A_3|A_1\cap A_2)\ldots P(A_n|A_1\cap A_2\cap \cdots \cap A_{n-1}) \end{aligned} \quad \cdots \quad (7.35)$$

এই ঘটনাগুলিকে যৌথভাবে পরস্পর সম্ভাবনাসূত্রে অনধীন বলা হয় যদি নিম্নলিখিত প্রতিটি $(2^n - n - 1)$ সংখ্যক সর্ত একত্রে খাটে। সর্তগুলি হচ্ছে

$$\left.\begin{aligned} &\text{প্রত্যেক } i, j\ (i < j) = 1, 2, \ldots n\text{-এর জন্যে}\\ &\qquad P(A_i\cap A_j) = P(A_i)\,P(A_j),\\ &\text{প্রত্যেক } i, j, k\ (i < j < k) = 1, 2, \cdots n\text{-এর জন্যে}\\ &\qquad P(A_i\cap A_j\cap A_k) = P(A_i)P(A_j)P(A_k)\\ &\qquad \vdots\\ &\text{এবং } P(A_1\cap A_2\cap \cdots \cap A_n) = P(A_1)\,P(A_2)\cdots PA_n). \end{aligned}\right\} \quad (7.36)$$

এখন আমরা একটি উদাহরণ নিয়ে দেখাব যে, তিনটি ঘটনার প্রতি দুটি ঘটনা পরস্পর অনধীন হলেও যৌথভাবে তারা অনধীন না হতে পারে।

একটি মুদ্রা দুবার উৎক্ষিপ্ত হলে অবেক্ষণযোগ্য মৌলিক ঘটনাগুলি হবে $HH$, $HT$, $TH$ এবং $TT$ ( অর্থাৎ দুটিতেই সম্মুখ, প্রথমটিতে সম্মুখ ও দ্বিতীয়টিতে পশ্চাৎ পার্শ্ব ইত্যাদি )। ধর, তিনটি ঘটনা হচ্ছে,

$$A_1=\{HH, HT\}, A_2=\{HH, TH\} \text{ ও } A_3=\{HH, TT\}.$$

তাহলে, $A_1 \cap A_2 = A_1 \cap A_3 = A_2 \cap A_3 = A_1 \cap A_2 \cap A_3 = \{HH\}$.

আবার, $P(A_1)=P(A_2)=P(A_3)=\frac{1}{2}$,

$P(A_1 \cap A_2)=P(A_2 \cap A_3)=P(A_1 \cap A_3)=P(A_1 \cap A_2 \cap A_3)=\frac{1}{4}$.

সুতরাং $P(A_i \cap A_j)=P(A_i)P(A_j)$, $i \neq j=1, 2, 3$ অর্থাৎ প্রত্যেক জোড়া ঘটনাই পরস্পর অনধীন। কিন্তু $P(A_1)\,P(A_2)\,P(A_3)=(\frac{1}{2})^3 \neq \frac{1}{4}=P(A_1 \cap A_2 \cap A_3)$, অর্থাৎ ঘটনাগুলি যৌথভাবে পরস্পর অনধীন নয়।

## 7.8 কয়েকটি উদাহরণ :

**উদা 7.11** একটি মুদ্রা ও একটি ছক্কা পর্যায়ক্রমে বারবার নিক্ষিপ্ত হলে ছক্কায় প্রথমবার 6 নির্দেশক চিহ্ন দৃষ্ট হওয়ার আগে মুদ্রায় সম্মুখপার্শ্ব দৃষ্ট হওয়ার সম্ভাবনা কত?

ছক্কায় 6 সূচক চিহ্নের আগে মুদ্রাটিতে প্রত্যেকটি নিক্ষেপণে পশ্চাৎপার্শ্ব দৃষ্ট হওয়ার ঘটনাকে $E$ সংকেতসূত্রে প্রকাশ করলে আলোচ্য ঘটনাটি হবে তার পরিপূরক $E^*$. এখন, $E$ ঘটনাটি কয়েকটি পরস্পরব্যতিরেকী রূপে ঘটতে পারে ; যথা :

1. প্রথম নিক্ষেপণে ছক্কায় 6 পড়বে ও তার আগে মুদ্রায় পশ্চাৎপার্শ্ব দেখা যাবে,

2. প্রথম নিক্ষেপণে ছক্কায় 6 পড়বে না কিন্তু দ্বিতীয়বার 6 পড়বে এবং ইতিমধ্যে মুদ্রায় দু'বারই পশ্চাৎপার্শ্ব দেখা যাবে,

3. প্রথম দুটি নিক্ষেপণে ছক্কায় 6 পড়বে না, কিন্তু তৃতীয় নিক্ষেপণে 6 পড়বে এবং ইতিমধ্যে মুদ্রায় তিনবারই কেবল পশ্চাৎপার্শ্ব দেখা যাবে, ইত্যাদি এবং একাদিক্রমে অসংখ্যবার এইরকম হতে থাকবে।

তাহলে $r$-তম $(r=1, 2, \cdots)$ নিক্ষেপণে মুদ্রায় পশ্চাৎপার্শের আবির্ভাব $A_r$, এবং ছক্কায় 6 সূচক চিহ্নের আবির্ভাবকে $B_r$ এবং অন্য সংখ্যাসূচক চিহ্নের

আবির্ভাবকে $B_r^*$ দ্বারা নির্দেশ করলে সহজবোধ্য সংকেতসূত্র ব্যবহার ক'রে লিখতে পারি

$$E=\{A_1B_1\}+\{A_1B_1^*A_2B_2\}+\{A_1B_1^*A_2B_2^*A_3B_3\} \\ +\{A_1B_1^*A_2B_2^*A_3B_3^*A_4B_4\}+\cdots$$

তাহলে উপপাদ্য 1 থেকে পাই

$$P(E)=P(A_1B_1)+P(A_1B_1^*A_2B_2)+P(A_1B_1^*A_2B_2^*A_3B_3)+\cdots$$

এখন ছক্কায় এবং মুদ্রায় যে কোন ফল (outcome) দর্শাবার ঘটনা স্পষ্টতঃই পরস্পর অনধীন। কাজেই সম্ভাবনাতাত্ত্বিক অনধীন ঘটনার সংজ্ঞা ব্যবহার ক'রে ও উপপাদ্য 4 প্রয়োগ ক'রে পাই

$$P(E)=\tfrac{1}{2}\times\tfrac{1}{6}+\tfrac{1}{2}\times\tfrac{5}{6}\times\tfrac{1}{2}\times\tfrac{1}{6}+\tfrac{1}{2}\times\tfrac{5}{6}\times\tfrac{1}{2}\times\tfrac{5}{6}\times\tfrac{1}{2}\times\tfrac{1}{6}+\cdots$$

সুতরাং আমাদের নির্ণেয় সম্ভাবনা হচ্ছে

$$P(E^*)=1-P(E)=1-\left[\frac{1}{12}+\frac{1}{12}\cdot\frac{5}{12}+\frac{1}{12}\cdot\left(\frac{5}{12}\right)^2+\cdots\right]$$

$$=1-\frac{1}{12}\left[1+\frac{5}{12}+\left(\frac{5}{12}\right)^2+\cdots\right]=1-\frac{1}{12}\times\frac{1}{1-\frac{5}{12}}$$

$$=1-\tfrac{1}{7}=\tfrac{6}{7}.$$

**উদা. 7.12** ধরা যাক, দুটি পাত্রের প্রথমটিতে 3টি সাদা ও 2টি কালো এবং দ্বিতীয়টিতে 3টি সাদা, 1টি কালো এবং 2টি লাল বল রয়েছে। এখন প্রথম পাত্র থেকে সমসম্ভব উপায়ে একটি বল তুলে নিয়ে অপরটিতে রাখবার পর দ্বিতীয় পাত্র থেকে সমসম্ভব পদ্ধতিতে একটি বল তোলা হলে সেটি সাদা হবার সম্ভাবনা কত?

ধরা যাক, প্রথম পাত্র থেকে সাদা ও কালো বল তোলার ঘটনাকে যথাক্রমে $A_1$ ও $B_1$ এবং দ্বিতীয় পাত্র থেকে সাদা বল উত্থিত হওয়ার ঘটনাকে $A_2$ চিহ্নে নির্দেশ করা হ'ল। এখন, $A_2$ ঘটনাটি দুটি পরস্পরব্যতিরেকী রূপে ঘটতে পারে; যথা: (1) প্রথম পাত্র থেকে সাদা বল তুলে সেটিকে দ্বিতীয় পাত্রে রাখবার পর দ্বিতীয় পাত্র থেকে তোলা বলটির রঙ সাদা হতে পারে অথবা (2) প্রথম পাত্র থেকে একটি কালো বল তুলে দ্বিতীয় পাত্রে রাখবার পর দ্বিতীয় পাত্র থেকে তোলা বলটির রঙ সাদা হতে পারে। তাহলে, সংকেতসূত্র ব্যবহার ক'রে লেখা যায়

$$A_2=(A_1\cap A_2)+(B_1\cap A_2).$$

তাহলে উপপাদ্য 1 অনুসারে, $P(A_2)=P(A_1\cap A_2)+P(B_1\cap A_2)$ এবং এবং উপপাদ্য 4 অনুসারে, $P(A_2)=P(A_1)P(A_2|A_1)+P(B_1)P(A_2|B_1)$;

এখানে $A_2|A_1$ হচ্ছে $A_1$-এর প্রাক্‌সংঘটন সর্তাধীনে $A_2$-এর সংঘটন এবং $A_2|B_1$ হচ্ছে $B_1$-এর প্রাক্‌সংঘটন সর্তাধীনে $A_2$ এর সংঘটন। এখন আমরা ধ'রে নেব যে পাত্রস্থিত বলগুলি সব সম-আকৃতিবিশিষ্ট এবং রঙ ছাড়া অন্য-সর্বপ্রকারে তারা অভিন্ন। কাজেই আমরা ধ'রে নিতে পারি যে এখানে একটি সুষম পরীক্ষণের ব্যাপার রয়েছে। কাজেই সহজেই পাওয়া যায়

$$P(A_1)=\tfrac{3}{5},\ P(B_1)=\tfrac{2}{5},\ P(A_2|A_1)=\tfrac{4}{7},\ P(A_2|B_1)=\tfrac{3}{7}.$$

সুতরাং নির্ণেয় সম্ভাবনা হচ্ছে $P(A_2)=\frac{3}{5}\times\frac{4}{7}+\frac{2}{5}\times\frac{3}{7}=\frac{18}{35}$.

**উদা. 7.13** একটি পুরস্কার জিতবার জন্যে দুজন খেলোয়াড় $A$ এবং $B$ খেলতে নামে। এই খেলায় স্থির হয় যে প্রথমে $A$ একটি ছক্‌কা নিক্ষেপ করবে; তাতে যদি 6-সূচক চিহ্ন ওঠে তবে $A$ জিতবে। সে না পারলে $B$ ছক্‌কা নিক্ষেপ করবে এবং তাতে সে যদি 5 বা 6-সূচক চিহ্ন পায় তবে সে-ই জিতবে। সে যদি না পারে, তবে $A$ আবার ছক্‌কা নিক্ষেপ করবে এবং যদি তাতে 4 বা 5 বা 6-সূচক চিহ্ন পায় তবে সে জিতবে। সে যদি না পারে তবে $B$ আবার ছক্‌কা নিক্ষেপ করবে এবং এইভাবে খেলাটি চলবে। তাহলে উভয় খেলোয়াড়ের পুরস্কার জয়ের সম্ভাবনা কত?

নিম্নোক্ত পরস্পরব্যতিরেকী ঘটনাগুলি ঘটলে $A$ জয়ী হবে; যথা:—

(1) প্রথম নিক্ষেপে $A$ 6-সূচক চিহ্ন পাবে, (2) প্রথম নিক্ষেপে $A$ 6-সূচক চিহ্ন পাবে না, $B$ তার প্রথম নিক্ষেপে 6 বা 5-সূচক চিহ্ন পাবে না এবং দ্বিতীয় নিক্ষেপে $A$ পাবে 6, 5 বা 4-সূচক চিহ্ন, (3) প্রথম নিক্ষেপে $A$ 6-সূচক চিহ্ন পাবে না, $B$ প্রথম নিক্ষেপে 6 বা 5-সূচক চিহ্ন পাবে না, দ্বিতীয় নিক্ষেপে $A$ 6, 5 বা 4-সূচকটি পাবে না, $B$ তার দ্বিতীয় নিক্ষেপে 6, 5, 4 বা 3-সূচক চিহ্ন পাবে না এবং $A$ তার তৃতীয় নিক্ষেপে 6 বা 5 বা 4 বা 3 বা 2-সূচক চিহ্ন পাবে।

এখন, বিভিন্ন নিক্ষেপে ছক্‌কার মাথায় $i$ বা $j$ বা ... $(i, j=1, 2, \ldots, 6)$ সংখ্যাসূচক চিহ্নের আবির্ভাব ও তার বিপরীত ঘটনা যথাক্রমে $E_{i, j, \ldots}$ এবং $E^*_{i, j, \ldots}$ সংকেত চিহ্ন সাহায্যে প্রকাশ করব। তাহলে $E$ যদি $A$-এর জয়লাভের ঘটনা নির্দেশ করে, তবে আমরা লিখতে পারি

$$E=E_6+(E^*_6\cap E^*_{6,5}\cap E_{6,5,4})$$
$$+(E^*_6\cap E^*_{6,5}\cap E_{6,5,4}\cap E^*_{6,5,4,3}\cap E_{6,5,4,3,2})$$

এখন লক্ষণীয় যে ছক্কাটি নিক্ষেপ করার ফলে $A$ যে ফল পাচ্ছে তা $B$-এর কোন ফলপ্রাপ্তিকে প্রভাবিত করছে না। অর্থাৎ আমরা ধ'রে নিতে পারি, যে কোন ছক্কা নিক্ষেপণের সূত্রে $A$ বা $B$-এর যে কোন ফলপ্রাপ্তির ঘটনা অপর কোন নিক্ষেপণে তাদের যে কোন ফলপ্রাপ্তির ঘটনার অনধীন।

কাজেই অনধীন ঘটনার সংজ্ঞা এবং সামগ্রিক সম্ভাবনা উপপাদ্য ব্যবহার ক'রে পাওয়া যায়

$$P(E)=P(E_6)+P(E^*_6).P(E^*_{6,5}).P(E_{6,5,4})$$
$$+P(E^*_6).P(E^*_{6,5}).P.(E^*_{6,5,4}).P(E^*_{6,5,4,3}).P(E_{6,5,4,3,2})$$
$$=\tfrac{1}{6}+\tfrac{5}{6}\times\tfrac{4}{6}\times\tfrac{3}{6}+\tfrac{5}{6}\times\tfrac{4}{6}\times\tfrac{3}{6}\times\tfrac{2}{6}\times\tfrac{5}{6}=\tfrac{1}{6}+\tfrac{5}{18}+\tfrac{25}{324}=\tfrac{169}{324}.$$

তেমনিভাবে $B$-এর জয়লাভের সম্ভাবনাও সরাসরি নির্ণয় করা যায়। কিন্তু যেহেতু $B$ জয়ী হওয়ার ঘটনা $A$ জয়ী হওয়ার ঘটনার পরিপূরক কাজেই $B$ জয়ী হওয়ার সম্ভাবনা হচ্ছে $1-P(E)=1-\frac{169}{324}=\frac{155}{324}$.

**উদা. 7.14** একজন খেলোয়াড় $A$ অপর দুজন প্রতিযোগী $B$ ও $C$-এর বিরুদ্ধে খেলে যদি উপর্যুপরি অন্ততঃ দুটি খেলায় জিততে পারে তবে সে একটি পুরস্কার পেতে পারে। $B$ ও $C$-এর বিরুদ্ধে প্রতি খেলায় $A$ জয়ী হওয়ার সম্ভাবনা যথাক্রমে $p$ ও $q$ এবং $p>q$. তাকে যদি (1) প্রথমে $B$, তারপর $C$ এবং সবশেষে $B$ অথবা (2) প্রথমে $C$, তারপর $B$ এবং সবশেষে $C$-এর সঙ্গে খেলবার সুযোগ দেওয়া হয়, তাহলে কোন্ পর্যায়ক্রমে খেললে তার বেশী সুবিধে হবে?

$A$ যদি প্রথমে $B$, তারপর $C$ ও সবশেষে $B$-এর বিরুদ্ধে খেলে তাহলে সে পুরস্কার জিতবে; যদি ($a$) প্রতিটি খেলায় জেতে অথবা ($b$) প্রথম দুটি খেলায় জয়ী হয়ে তৃতীয় খেলায় পরাজিত হয় অথবা ($c$) প্রথম খেলায় পরাজিত হয়ে বাকী দুটি খেলায় পরপর জয়ী হয়। তাহলে, যেহেতু স্পষ্টতঃই প্রতি খেলায় বিজয়ী বা বিজিত হওয়ার ঘটনাগুলি সব পরস্পর অনধীন, তাই এক্ষেত্রে $A$ জয়ী হওয়ার সম্ভাবনা দাঁড়ায়

$$P_1=pqp+pq(1-p)+(1-p)qp=pq(p+1-p+1-p)=pq(2-p)$$

[ এক্ষেত্রে $A$ জয়ী হওয়ার সম্ভাবনাকে $P_1$ দ্বারা নির্দেশ ক'রে সামগ্রিক এবং মিশ্র-সম্ভাবনা উপপাদ্য ব্যবহার করা হয়েছে। ]

পক্ষান্তরে, প্রথমে $C$, তারপর $B$ ও সবশেষে $C$-এর বিরুদ্ধে খেললে একই রকম যুক্তিতে সেক্ষেত্রে $A$ জয়ী হওয়ার সম্ভাবনা $P_2$-এর মান পাওয়া যাবে

$$P_2 = qpq + qp(1-q) + (1-q)pq = pq\{q + (1-q) + (1-q)\}$$
$$= pq(2-q).$$

এখন, $P_1 - P_2 = pq(q-p) < 0$ যেহেতু $p > q$.

সুতরাং, $P_1 < P_2$ কাজেই দ্বিতীয় পর্যায়ক্রমে খেলা $A$-এর পক্ষে বেশী সুবিধাজনক।

## 7.9 পুরাতনী সম্ভাবনাতত্ত্বের দোষত্রুটি :

সম্ভাবনার পুরাতনী তত্ত্বের কয়েকটি ত্রুটি আছে। যেমন, প্রথমতঃ, পরীক্ষণ-সংশ্লিষ্ট মৌলিক ঘটনাবলীর সংখ্যা অগণিত হলে কোন ঘটনার সম্ভাবনার সংজ্ঞা নির্দেশ করা যাবে না। 'কোন নবজাত মানবশিশু পূর্ণবয়স্ক হবার পর তার দৈর্ঘ্য 5 ফুট থেকে 6 ফুটের মধ্যে থাকবে'—এ জাতীয় ঘটনার সম্ভাবনা কী তা সম্ভাবনার পুরাতনী সংজ্ঞার সাহায্যে নির্ণয় করা যাবে না। কারণ, এখানে পরীক্ষণের ফল অসীমসংখ্যক হতে পারে, কেননা কোন ব্যক্তির প্রকৃত দৈর্ঘ্যের মান অগণিত প্রকৃত রাশির (real number) যে কোন একটি হতে পারে।

দ্বিতীয়তঃ, পরীক্ষণের প্রকৃতিতে যদি সুষমতা না থাকে, তবে ওপরের সংজ্ঞা প্রযোজ্য নয়। একটি বিশেষভাবে তৈরী ছক্কার বিভিন্ন প্রান্ত যদি অসমভাবে ভারযুক্ত হয়, তবে সেটি নিক্ষিপ্ত হলে তার প্রান্তগুলি ছক্কার ওপরে থাকার মৌলিক ঘটনাবলীকে সমসম্ভব ব'লে ধরা ঠিক হবে না। এক্ষেত্রে ছক্কাটির সুষমতাগুণ থাকবে না, ফলে পরীক্ষণটিও সুষম হবে না। কাজেই এই ছক্কা-নিক্ষেপণের পরীক্ষণে কোন সংখ্যাসূচক চিহ্নই ছক্কার ওপরে দৃষ্ট হওয়ার সম্ভাবনা পুরাতনী সংজ্ঞানুযায়ী নির্ণয় করা যাবে না।

তৃতীয়তঃ, সম্ভাবনার পুরাতনী সংজ্ঞায় বৃত্তীয় যুক্তির প্রমাদও পরিলক্ষিত হয়। কারণ, এখানে সম্ভাবনা নির্ণয় করতে পরীক্ষণ-সংশ্লিষ্ট মৌলিক ঘটনাগুলিকে সমসম্ভব ব'লে ধরা হয় কিন্তু তার আগে সমসম্ভব বলতে ঠিক কী বোঝায় এই প্রশ্নটি মোটামুটি এড়িয়ে যাওয়া হয়।

পুরাতনী তত্ত্বে এই সমস্ত খুঁত রয়েছে ব'লে সম্ভাবনাতত্ত্বকে দৃঢ়তর ভিত্তির ওপর প্রতিষ্ঠিত করার সবিশেষ প্রয়োজন অনুভূত হয় এবং এসম্পর্কে প্রচুর আলোচনা ও গবেষণা হয়। ফলে সম্ভাবনার ভিন্নতর তত্ত্বের উদ্ভব হয়েছে এবং

তন্মধ্যে **স্বীকার্যভিত্তিক তত্ত্বই** (axiometic theory) বর্তমানে সাধারণভাবে সবচেয়ে বেশী স্বীকৃতি লাভ করেছে। এ সম্পর্কে বিস্তারিত আলোচনার অবকাশ আমাদের নেই। আমরা কেবল সংক্ষেপে দু-একটি কথা বলব।

সম্ভাবনাভিত্তিক পরীক্ষণ সম্পর্কে এটা সাধারণতঃ স্বীকার করা হয় যে, যদি পারিপার্শ্বিক পরিস্থিতিগুলি সর্বদা অন্ততঃ কার্যতঃ অবিকৃত থাকে, তবে এর যথেচ্ছসংখ্যক পুনরনুষ্ঠান সম্ভব এবং এ অবস্থায় পরীক্ষণ-সংশ্লিষ্ট ঘটনাপুঞ্জের স্বরূপ-প্রকৃতি ও তাদের পারস্পরিক সম্পর্ক অপরিবর্তিত থাকে। এই স্বীকরণ-সাপেক্ষে সাধারণতঃ দেখা যায় যে, পরীক্ষণটি $n$-সংখ্যক বার অনুষ্ঠিত হলে তাতে সংঘটিত কোন ঘটনা $A$-এর পরিসংখ্যা যদি $f_A$ হয়, তবে $n$-এর মান যতই বাড়তে থাকে, বিভিন্ন $n$-এর জন্যে $\frac{f_A}{n}$ অনুপাতটির মানের পার্থক্য ততই কমতে থাকে এবং এর মান ক্রমেই একটি সীমামানের (limiting value) অভিমুখে অগ্রসর হয়। স্বীকার্যভিত্তিক সম্ভাবনাতত্ত্বে এই সীমামানটিকেই $A$ ঘটনার সম্ভাবনা ব'লে ধরা হয় যদিও এই সীমার মান ঠিক কত তা নির্দেশ করার চেষ্টা করা হয় না। বাস্তবিক, এই তত্ত্বে কোন ঘটনার সম্ভাবনার মান ধ্রুবক হিসেবে নির্দিষ্ট করা হয় না। কিন্তু পরীক্ষণে অবেক্ষিত কোন ঘটনার আপেক্ষিক পরিসংখ্যাকে তার সম্ভাবনার একটি আসন্নমান হিসেবে দেখা হয়। কোন ঘটনা $A$-এর সম্ভাবনাকে $P(A)$ সংকেতসূত্রে নির্দেশ করলে তার বাস্তবিক মান যতই হোক, অবেক্ষিত আপেক্ষিক পরিসংখ্যা $\frac{f_A}{n}$-এর মানের প্রকৃতির ভিত্তিতে $P(A)$-এর মান সম্পর্কে কয়েকটি সাধারণ প্রতিজ্ঞা স্বীকার ক'রে নেওয়া যায়। যেমন,

সব $A$-এর জন্যেই

$$P(A) \geqslant 0 \qquad \cdots \qquad \cdots \quad (7.37)$$

$A$ ও $B$ দুটি পরস্পরব্যতিরেকী ঘটনা হলে

$$P(A \cup B) = P(A) + P(B), \qquad \cdots \qquad \cdots \quad (7.38)$$

$A_1, A_2, A_3, \ldots$ সকলে যৌথভাবে পরস্পরব্যতিরেকী হলে

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + (A_3) + \cdots, \qquad \cdots \quad (7.39)$$

যে কোন ঘটনাদ্বয় $A$ ও $B$-এর জন্যে

$$P(B) > 0 \text{ হলে } P(A \cap B) = P(B)P(A|B) \quad \cdots \quad (7.40)$$

এবং $$P(\Omega) = 1 \quad \cdots \quad (7.41)$$

—এই সম্পর্কগুলিকে স্বতঃসিদ্ধ হিসেবে গ্রহণ ক'রে এদের থেকে কতগুলি উপপাদ্য ও অনুসিদ্ধান্ত ইত্যাদি প্রমাণ ক'রে স্বীকার্যভিত্তিক সম্ভাবনাতত্ত্বের প্রতিষ্ঠা হয়েছে। এখানে উল্লেখযোগ্য যে, দেখা গেছে যে এই স্বীকরণগুলি ও তাদের থেকে অনুসৃত উপপাদ্যগুলি এবং পুরাতনী তত্ত্বের উপপাদ্য ও অনুসিদ্ধান্ত ইত্যাদি পরস্পরবিরোধী নয়।

## 7.10 জ্যামিতিক সম্ভাবনা (geometric probability) :

পরীক্ষণ-সংশ্লিষ্ট মৌলিক ঘটনাবলীর সংখ্যা সসীম হতে হবে এই বাধ্যবাধকতার জন্যে ব্যবহারিক ক্ষেত্রে সম্ভাবনার পুরাতনী সংজ্ঞার প্রয়োগ যে সীমিত হয়ে পড়ে সে সম্পর্কে প্রাচীন সম্ভাবনাতাত্ত্বিকগণও অবহিত ছিলেন। একটি বিশেষ ক্ষেত্রে এই অসুবিধে দূর ক'রে সম্ভাবনা সংজ্ঞা কিছু প্রসারিত করার চেষ্টাও বহুদিন আগেই হয়েছিল। কোন প্রদত্ত বৃত্ত, চতুর্ভুজক্ষেত্র, গোলক বা সরলরেখা ইত্যাদি জ্যামিতিক চিত্রসত্তার অভ্যন্তরে যদি কোন বিন্দু পক্ষপাতিত্বহীনভাবে বেছে নেওয়া হয়, তাহলে একটি সম্ভাবনাসাপেক্ষ পরীক্ষণের ব্যাপারে ঘটেছে ব'লে স্বীকার করা যায়। অনেক সময় গৃহীত বিন্দুটি প্রদত্ত জ্যামিতিক ক্ষেত্রটির মধ্যবর্তী কোন বিশেষ অঞ্চলভুক্ত হওয়ার সম্ভাবনা জানতে আমাদের আগ্রহ হয়। এখানে বিন্দুসংখ্যা অর্থাৎ মৌলিক ঘটনার সংখ্যা স্পষ্টতঃই অসীম। কাজেই সম্ভাবনার পুরাতনী সংজ্ঞা ব্যবহারযোগ্য নয়। তাই বিকল্প সংজ্ঞার প্রয়োজন। এই উদ্দেশ্যে প্রচলিত রীতিটি নিম্নরূপ: পরীক্ষণের ফলস্বরূপ যে $W$ ক্ষেত্রটির মধ্যে বিন্দুটি গৃহীত হবে প্রথমে তার একটি বিশেষ পরিমাপ $M(W)$ স্থির করা হবে। এখানে $M$ হচ্ছে একটি প্রকৃত রাশিভিত্তিক (real-valued) অপেক্ষক যা $W$-এর প্রত্যেক অংশের জন্যে নির্দিষ্ট মান নেবে এবং ঐ অংশগুলির ক্ষেত্রফলের পরিমাপ বৃদ্ধির সঙ্গে সঙ্গে তাদের জন্যে $M$-এর মানেরও ক্রমাগত বৃদ্ধি হবে। এখন $\omega$ যদি $W$-এর অন্তর্গত একটি অঞ্চল হয় তাহলে স্বভাবতঃই $M(\omega) \leq M(W)$ হবে এবং পরীক্ষণসূত্রে গৃহীত বিন্দুটি $\omega$-এর মধ্যবর্তী হবার সম্ভাবনার মান

$$P(\omega) = \frac{M(\omega)}{M(W)} \quad \cdots \quad (7.42)$$

ব'লে ধরা হবে। সম্ভাবনার এই সংজ্ঞাকে জ্যামিতিক সম্ভাবনা বলা হয়। এই সংজ্ঞার প্রয়োগে অনেক সময় সমাকলন (integration)-এর সাহায্যে বিভিন্ন জ্যামিতিক ক্ষেত্রের পরিমাপ নির্ণয় করতে হয় এবং ঐ সমাকলনে সাধারণতঃ এক বা একাধিক অবিচ্ছিন্ন চলের অবতারণা করতে হয়। এই প্রসঙ্গে এখন আমরা কয়েকটি উদাহরণ নিয়ে আলোচনা করব।

## 7.11 জ্যামিতিক সম্ভাবনা সম্পর্কে কয়েকটি উদাহরণ:

**উদা. 7.15** মধ্যবিন্দু $O$ এবং দৈর্ঘ্য $l$ বিশিষ্ট একটি ঋজুরৈখিক ক্ষেত্র $AB$-এর মধ্যে সমসম্ভব উপায়ে একটি বিন্দু $X$ নেওয়া হলে $AX$, $BX$ এবং $AO$ এই তিনটি ঋজুরৈখিক অংশ একত্রে একটি ত্রিভুজ গঠন করার সম্ভাবনা কত?

আমরা জানি যে, $AX$, $BX$ ও $AO$ অংশত্রয় একত্রযোগে একটি ত্রিভুজ গঠন করতে হলে নিম্নলিখিত সর্তাবলীর অন্ততঃ একটিকে খাটতে হবেই; যথা:—

(1) $AX+BX > AO$

(2) $AX+AO > BX$

(3) $BX+AO > AX$

এখন, $X$ যদি $A$ এবং $O$-এর মধ্যে থাকে, তাহলে

$$BX=BO+OX=AO+OX$$

সুতরাং আমাদের দরকার $AX+AO > BX$

অর্থাৎ, $AX+AO > AO+OX$

অর্থাৎ, $AX > OX$.

এক্ষেত্রে $BX+AO > AX$ এবং $AX+BX > AO$

এই সর্ত-দুটি স্পষ্টতঃই খাটে। কাজেই, $C$ যদি $AO$-এর মধ্যবিন্দু হয়, তবে $X$, $C$ এবং $O$-এর মধ্যে থাকবে। তেমনি, $D$ যদি $OB$-এর মধ্যবিন্দু হয়, তবে $X$-কে অবশ্যই $O$ এবং $D$-এর মধ্যে থাকতে হবে যদি (1)—(3) সর্তাবলীর অন্ততঃ একটি পালিত হতে হয়। কাজেই, প্রশ্নে উল্লিখিত সর্ত মানতে গেলে

নির্বাচিত বিন্দু $X$-কে $AB$ রেখার $CD$ অংশমধ্যে থাকতে হবে। তাহলে সম্ভাবনার সংজ্ঞা অনুযায়ী বলা যেতে পারে যে নির্ণেয় সম্ভাবনা হচ্ছে

$$\frac{CD \text{ রেখার দৈর্ঘ্য}}{AB \text{ রেখার দৈর্ঘ্য}} = \frac{\frac{l}{2}}{l} = \frac{1}{2}.$$

এখানে বলা বাহুল্য যে জ্যামিতিক ক্ষেত্রের পরিমাপ হিসেবে রেখার দৈর্ঘ্যকে নেওয়া হয়েছে।

**উদা. 7.16** একটি ঋজুরৈখিক ক্ষেত্রে সমসম্ভব উপায়ে তিনটি বিন্দু $X_1$, $X_2$ এবং $X_3$ নির্বাচিত হলে $X_3$ যে $X_1$ ও $X_2$-এর মধ্যে থাকবে তার সম্ভাবনা কত ?

ধরা যাক, রেখাটি হচ্ছে $AB$ এবং তার বামপ্রান্ত $A$ থেকে $X_1$, $X_2$ ও $X_3$-এর দূরত্ব হচ্ছে যথাক্রমে $x_1$, $x_2$ ও $x_3$.

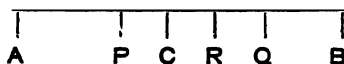
তাহলে নিম্নলিখিত ছটি বিকল্প পরস্পরব্যতিরেকী ও পরস্পরনিঃশেষী মৌলিক ঘটনা ঘটতে পারে এবং বিন্দু-তিনটি পক্ষপাতিত্বহীনভাবে নেওয়া ব'লে ধরলে এদেরকে সমসম্ভব ব'লেও স্বীকার করা যায়। এগুলি হচ্ছে

(1) $x_1 \leqslant x_2 \leqslant x_3$, (2) $x_1 \leqslant x_3 \leqslant x_2$, (3) $x_2 \leqslant x_3 \leqslant x_1$,

(4) $x_2 \leqslant x_1 \leqslant x_3$, (5) $x_3 \leqslant x_1 \leqslant x_2$ এবং (6) $x_3 \leqslant x_2 \leqslant x_1$.

এই ছটির মধ্যে দুটি অর্থাৎ (2) ও (3) নম্বর মৌলিক ঘটনা হচ্ছে প্রশ্ননির্দিষ্ট ঘটনাটির অনুকূল। সুতরাং এক্ষেত্রে পুরাতনী সংজ্ঞা প্রযোজ্য এবং নির্ণেয় সম্ভাবনার মান হচ্ছে $\frac{2}{6} = \frac{1}{3}$.

**উদা. 7.17** নির্দিষ্ট দৈর্ঘ্যযুক্ত একটি রেখার ওপর সমসম্ভব উপায়ে দুটি বিন্দু নিয়ে তাকে তিনটি ভাগে ভাগ করলে তিনটি অংশ দিয়ে একটি ত্রিভুজ তৈরী করা যাবে এমন সম্ভাবনা কত ?

A P C R Q B

ধরা যাক, প্রদত্ত $AB$ রেখার দৈর্ঘ্য $a$ এবং তার মধ্যবিন্দু $C$ ও তার ওপর সমসম্ভব উপায়ে দুটি বিন্দু $P$ ও $R$ নেওয়া হয়েছে।

প্রথম ক্ষেত্র : $AP < \frac{a}{2}$.

লেখা যাক $AP = x$ এবং $BP = a - x$.

মনে কর, $P$ বিন্দুটি এমনভাবে নেওয়া হয়েছে যে, $AB$ রেখার ওপর যে কোন দৈর্ঘ্যবিশিষ্ট অন্তরে এটি নির্বাচিত হওয়ার সম্ভাবনা সমান এবং এটি যে কোন ক্ষুদ্র $dx$ দৈর্ঘ্যবিশিষ্ট অন্তরমধ্যে পড়বার সম্ভাবনা হচ্ছে $\frac{dx}{a}$. ধরা যাক, $Q$ হচ্ছে $AB$-এর ওপর কোন বিন্দু যার জন্যে $PQ=\frac{a}{2}$. তাহলে, $R$ বিন্দুটি যদি এমনভাবে নির্বাচিত হয় যে প্রশ্ননির্দিষ্ট সর্তটি খাটে, তাহলে $R$-কে $C$ ও $Q$-এর মধ্যে থাকতে হবে। কারণ, অন্যথায় $PR > AP+BR$ হবে এবং প্রশ্নের সর্তটি খাটবে না। তেমনি $R$ বিন্দু $P$ ও $C$-এর মধ্যেও থাকতে পারে না, কারণ তাহলে $BR > AP+RP$ হবে ও প্রশ্নের সর্তটি খাটবে না। কাজেই $R$-কে অবশ্যই $C$ ও $Q$-এর মধ্যে থাকতে হবে। কাজেই, $P$ বিন্দু $x$ এবং $x+dx$-এর মধ্যে থাকবে, $x<\frac{a}{2}$ হবে এবং $R$ এমনভাবে অবস্থিত হবে যে, $AP$, $PR$ ও $RB$ অংশত্রয় এমন হবে যে তাদের যে কোন দুটির সমষ্টি তৃতীয়টির চেয়ে বড় হবে এমন ঘটনার সম্ভাবনা হবে

$$\frac{CQ}{AB}\times\frac{dx}{a}.$$

সুতরাং এক্ষেত্রে, অর্থাৎ যখন $x<\frac{a}{2}$, নির্ণেয় সম্ভাবনা হচ্ছে

$$P_1=\int_0^{\frac{a}{2}}\frac{CQ}{a}\cdot\frac{dx}{a}=\frac{1}{a^2}\int_0^{\frac{a}{2}}x\,dx$$

$$\left[\text{কারণ, } CQ=AQ-AC=AP+PQ-AC=x+\frac{a}{2}-\frac{a}{2}=x\right]$$

$$=\frac{1}{a^2}\left[\frac{x^2}{2}\right]_0^{\frac{a}{2}}=\frac{1}{a^2}\cdot\frac{1}{2}\cdot\frac{a^2}{4}=\frac{1}{8}.$$

দ্বিতীয় ক্ষেত্র : $x>\frac{a}{2}$

এখানেও একইরকম যুক্তিসাহায্যে পাওয়া যায় যে, $P$ বিন্দু $x$ থেকে $x+dx$-এর মধ্যে $\left(x>\frac{a}{2}\right)$ এবং $R$ বিন্দু $AB$ রেখায় এমনভাবে অবস্থিত

হবে যে, $AP$, $PR$ ও $RB$-এর কোন অংশই অপর দুই অংশের সমষ্টির চেয়ে বড় হবে না। এই ঘটনার সম্ভাবনা হচ্ছে

$$\frac{CQ}{a}.\frac{dx}{a}.$$

সুতরাং এক্ষেত্রে, অর্থাৎ যখন $x > \frac{a}{2}$, নির্ণেয় সম্ভাবনা হচ্ছে

$$P_2 = \int_{\frac{a}{2}}^{a} \frac{CQ}{a}.\frac{dx}{a} = \frac{1}{a^2}\int_{\frac{a}{2}}^{a} (a-x)dx$$

$\Big[$ কারণ, এখানে $CQ = PQ - PC = \frac{a}{2} - (AP - AC)$

$$= \frac{a}{2} - x + \frac{a}{2} = a - x \Big]$$

$$= \frac{1}{a}\Big[\, x \,\Big]_{\frac{a}{2}}^{a} - \frac{1}{a^2}\left[\frac{x^2}{2}\right]_{\frac{a}{2}}^{a} = \frac{1}{a}\cdot\frac{a}{2} - \frac{1}{2}\cdot\frac{1}{a^2}\left(a^2 - \frac{a^2}{4}\right)$$

$$= \tfrac{1}{8}.$$

সুতরাং, নির্ণেয় সম্ভাবনা হচ্ছে $P_1 + P_2 = \frac{1}{4}$.

**উদা. 7.18** $3a$ দৈর্ঘ্যবিশিষ্ট একটি সরলরেখা $AB$-এর ওপর সমসম্ভব উপায়ে একটি বিন্দু $P$ বেছে নিলে এবং তারপর $AP$ অংশে অন্য একটি বিন্দু $Q$ একইভাবে বেছে নিলে $PQ$-এর দৈর্ঘ্য $a$-এর চেয়ে বেশী হওয়ার সম্ভাবনা কত?

ধরা যাক, $AP = x$ এবং $QP = y$.

তাহলে $y > a$ হলে $x > a$ হবে। এখন, $x$ যদি নির্দিষ্ট থাকে তবে $Q$ যেহেতু সমসম্ভব উপায়ে গৃহীত হয়েছে $AP$-এর মধ্যবর্তী $dy$ দৈর্ঘ্যবিশিষ্ট যে কোন অন্তরে $Q$ অবস্থিত হবার সম্ভাবনা হবে $\frac{dy}{x}$. আবার, $3a$ দৈর্ঘ্যবিশিষ্ট $AB$ রেখার ওপর যে কোন $dx$ দৈর্ঘ্যবিশিষ্ট অন্তরমধ্যে $P$ বিন্দু থাকবার সম্ভাবনা হচ্ছে $\frac{dx}{3a}$. কাজেই নির্ণেয় সম্ভাবনা হচ্ছে

$$\int_0^{3a}\int_a^{x} \frac{dy}{x}.\frac{dx}{3a} = \frac{1}{3a}\int_a^{3a} \frac{1}{x}\Big[\, y \,\Big]_a^{x} dx$$

$$= \frac{1}{3a}\int_a^{3a} \frac{x-a}{x} dx = \frac{1}{3a}\Big[\, x - a \log_e x \Big]_a^{3a}$$

$$= \frac{1}{3a}\left[\, 2a - a\log_e 3a + a\log_e a \,\right]$$

$$- \frac{1}{3}\log_e\left(\frac{3a}{a}\right) = \frac{1}{3}\left[\, 2 - \log_e 3 \,\right].$$

## 7.12 সম্ভাবনাশ্রয়ী চল এবং গাণিতিক প্রত্যাশা (random variable and mathematical expectation) :

মনে কর, কোন সম্ভাবনাভিত্তিক পরীক্ষণের সঙ্গে সংশ্লিষ্ট মৌলিক ঘটনাগুলির বিভিন্নতা অনুযায়ী একটি প্রকৃতমানাশ্রয়ী চল $X$-কে বিভিন্ন মান আরোপ করা হবে। তাহলে, $X$-এর মান কোন প্রকৃতরাশির গুচ্ছের অন্তর্ভূত হওয়ার ব্যাপারটিকে একটি ঘটনা বলা যায়। বাস্তবিক, এই ঘটনা হচ্ছে যে সমস্ত মৌলিক ঘটনার জন্যে $X$ ঐরকম মান গ্রহণ করছে সেগুলি একত্রে যে ঘটনা নির্দেশ করে তার সঙ্গে অভিন্ন। এখন, এই ঘটনার যা সম্ভাবনা, $X$-এর মান ঐ প্রকার হবারও সেই সম্ভাবনা আছে ব'লে ধরা হয়। এক্ষেত্রে $X$-কে একটি **সম্ভাবনাশ্রয়ী চল** বা সংক্ষেপে **সম্ভাবনা চল** (random variable) বলা হয়। সংক্ষেপে, $X$-এর মান বিভিন্ন গুচ্ছের অন্তর্ভুক্ত হবার যদি নির্দিষ্ট সম্ভাবনা থাকে, তাহলেই $X$-কে সম্ভাবনা চল বলা হবে।

কোন পরীক্ষণ-সংশ্লিষ্ট মৌলিক ঘটনাবলীর বিভিন্নতা অনুযায়ী মনে কর একটি সম্ভাবনা চল $X$-এর মানগুলি একটি নির্দিষ্ট অন্তর $[a, b]$-এর মধ্যে থাকে কিনা তা বারবার অবেক্ষণ করা হতে থাকবে। পরীক্ষণটির বহুসংখ্যক পুনরনুষ্ঠানে যে অনুপাতে $X$-এর মান $[a, b]$ অন্তরে থাকবে তাকে $X$-এর মান $[a, b]$ অন্তরের মধ্যবর্তী হবার সম্ভাবনার একটি প্রাক্কলক হিসেবে সাধারণতঃ নেওয়া হয়। লক্ষণীয় যে, $[a, b]$ অন্তর যদি সমগ্র প্রকৃত রাশিমালার গুচ্ছ অর্থাৎ $(-\infty, \infty)$ অন্তরের সমান হয় তবে স্পষ্টতঃই পরীক্ষণের প্রতি অনুষ্ঠানেই $X$-এর মান $(-\infty, \infty)$-এর মধ্যে থাকবেই। $X$-এর মান $[a, b]$-এর মধ্যে থাকবার সম্ভাবনাকে $P[a \leqslant X \leqslant b]$ সংকেতসূত্রে প্রকাশ করলে নিম্নলিখিত সম্পর্কগুলি অবশ্যই খাটবে ব'লে স্বতঃসিদ্ধ হিসেবে ধরা যায় ; যথা :—

$a$ ও $b$-এর মান $(a \leqslant b)$ যাই হোক না কেন

$$P[a \leqslant X \leqslant b] \geqslant 0 \; ; \qquad \cdots \qquad \cdots \qquad (7.43)$$

যে কোন $a < b < c < d$-এর জন্যে

$$P[a < X < b, \text{ অথবা } c < X < d]$$
$$= P[a < X < b] + P[c < X < d] \quad \cdots \quad (7.44)$$

এবং $P[-\infty < X < \infty] = 1 \quad \cdots \quad \cdots \quad (7.45)$

একটি উদাহরণ নিয়ে সম্ভাবনা চলের ধর্ম একটু বিশদভাবে আলোচনা করা যাক। একটি সুসমঞ্জস মুদ্রা উৎক্ষেপণের পরীক্ষণে দুটি লক্ষণীয় ফলাফল যথা 'সম্মুখপার্শ্ব' ও 'পশ্চাৎপার্শ্ব' দৃষ্ট হলে যথাক্রমে ধরা যাক একটি চল $X$-কে 1 ও 0 মান আরোপ করা হবে। অর্থাৎ উৎক্ষিপ্ত মুদ্রায় যতবার সম্মুখপার্শ্ব দেখা যাবে $X$ হচ্ছে তারই সংখ্যা। তাহলে $X$-এর 1 মান গ্রহণ করা এবং মুদ্রায় সম্মুখপার্শ্ব দৃষ্ট হওয়া হচ্ছে একই ঘটনা। এই ঘটনাকে $[X=1]$ সংকেতসূত্রে প্রকাশ করলে $P[X=1] = P($ মুদ্রায় সম্মুখপার্শ্ব দৃষ্ট হওয়া $) = \frac{1}{2}$. তেমনি $P[X=0] = P($ মুদ্রায় পশ্চাৎপার্শ্ব দৃষ্ট হওয়া $) = \frac{1}{2}$. কাজেই $X$-এর 1 ও 0 এই উভয় মান গ্রহণ করার একটি ক'রে নির্দিষ্ট সম্ভাবনা রয়েছে। কাজেই এই $X$ চলটিকে একটি সম্ভাবনা চল বলা হবে এবং 1 ও 0-কে আমরা $X$-এর দুটি সম্ভাব্য মান ব'লে উল্লেখ করব।

সাধারণভাবে, কোন সম্ভাবনা চল $X$ যদি $x_1, x_2, \ldots x_n, \ldots$ ইত্যাদি কতগুলি বিচ্ছিন্ন মান গ্রহণ করে এবং প্রত্যেক $i = 1, 2, \ldots n, \ldots$-এর জন্যে $[X = x_i]$ ঘটনাটির নির্দিষ্ট সম্ভাবনা $P[X = x_i] = p_i > 0$ থাকে, এবং $\sum_{i=1}^{\infty} p_i = 1$ হয়, তবে $X$-কে একটি **বিচ্ছিন্ন সম্ভাবনা চল** (discrete random variable) বলা হয়। যেমন, একটি মুদ্রা যদি পরপর 10 বার উৎক্ষিপ্ত হয় এবং তাতে যতবার সম্মুখপার্শ্ব পাওয়া যাবে সেই সংখ্যা $X$ দ্বারা নির্দেশ করা হয়, তবে $X$ একটি সম্ভাবনা চল হবে এবং এর সম্ভাব্য মানগুলি হবে 0, 1, 2, ..., 9, 10. স্পষ্টতঃই $X$-এর মান এদের যে কোন একটি হবার এক একটি নির্দিষ্ট সম্ভাবনা রয়েছে।

পক্ষান্তরে, $X$ যদি এমন একটি সম্ভাবনা চল হয় যা কোন নির্দিষ্ট অন্তর $[a, b]$-এর মধ্যে অবিচ্ছিন্নভাবে যে কোন মান ধারণ করতে পারে এবং এর অন্তর্ভূত যে কোন উপ-অন্তরের (sub-interval) মধ্যে এর মান ধারণ করার নির্দিষ্ট সম্ভাবনা থাকে [ যদি $a < \alpha < \beta < b$ হয়, তবে $[\alpha, \beta]$-কে $[a, b]$-এর

একটি উপ-অন্তর বলা হবে ], তাহলে $X$-কে **অবিচ্ছিন্ন সম্ভাবনা চল** (continuous random variable) বলা হয়। এক্ষেত্রে কোন অন্তর $(\alpha, \beta)$-এর মধ্যে যদিও $X$ যে কোন মান ধারণ করতে পারে, তবুও ধরা হবে যে $X$ যে কোন একটি বিচ্ছিন্ন মান ( যেমন, ধর $x$) গ্রহণ করার সম্ভাবনা হচ্ছে শূন্য। অর্থাৎ ধরা হবে যে, $P[X=x]=0$, $x$ যাই হোক না কেন।

$X$ যদি কোন বিচ্ছিন্ন সম্ভাবনা চল হয় এবং $x$ তার কোন সম্ভাব্য মান হয়, তাহলে বলা হয় যে,

$$P[X=x]=f(x)$$

হচ্ছে $x$ বিন্দুতে গৃহীত $X$ চলের **সম্ভাবনা ভর অপেক্ষক** (probability mass function) $f$-এর মান। এই $f$ অপেক্ষকটি দেখায় পূর্ণসম্ভাবনা 1 কি-ভাবে $X$-এর বিভিন্ন $x$ মানগুলির মধ্যে নিবেশিত রয়েছে। এই জন্যে বলা হয় যে $f$ অপেক্ষকটি বিচ্ছিন্ন সম্ভাবনা চল $X$-এর সম্ভাবনা বিভাজন বা সম্ভাবনা নিবেশন (probability distribution) নির্দেশ করে।

এখন, যদি

$$\sum_x |x| f(x) < +\infty \qquad \cdots \qquad \cdots \quad (7.46)$$

হয়, তাহলে

$$\mu = \sum_x x f(x) = \sum_x x P[X=x] \qquad \cdots \quad (7.47)$$

কে বলা হয় $X$-এর **গাণিতিক প্রত্যাশা** (mathematical expectation)। এখানে $\sum_x$ দ্বারা $X$-এর সমস্ত সম্ভাব্য মান $x$-এর জন্যে সমষ্টি নির্দেশ করা হয়েছে। গাণিতিক প্রত্যাশা কে $E(X)$ বলেও উল্লেখ করা হয়। এছাড়া

$$\sigma^2 = \sum_x (x-\mu)^2 f(x) \qquad \cdots \qquad \cdots \quad (7.48)$$

কে বলা হয় $X$-এর ভেদমান এবং একে $E(X-\mu)^2$ দ্বারাও নির্দেশ করা হয়।

পক্ষান্তরে, $X$ যদি একটি অবিচ্ছিন্ন সম্ভাবনা চল হয়, তাহলে অনেক সময়ই একটি অপেক্ষক $f$-এর অস্তিত্ব থাকে যার বিশেষত্ব এই যে,

(1) $f$ একটি অ-ঋণাত্মক অবিচ্ছিন্ন অপেক্ষক অর্থাৎ প্রত্যেক $x$-এর জন্যে $f(x) \geqslant 0$

এবং (2) যে কোন অন্তর $(\alpha, \beta)$-এর জন্যে

$P[\alpha \leqslant X \leqslant \beta]$-কে $\int_{\alpha}^{\beta} f(x)\,dx$—এই সমাকলকের মাধ্যমে প্রকাশ করা যায়। এখানে $\alpha$ ও $\beta$ হচ্ছে $X$-এর মানসীমা $a$ ও $b$-এর মধ্যবর্তী যে কোন রাশি। বিশেষ উল্লেখযোগ্য যে, প্রত্যেক $\alpha$ ও $\beta$-এর জন্যেই

$P[\alpha \leqslant X \leqslant \beta] = P[\alpha < X \leqslant \beta] = P[\alpha < X < \beta] = P[\alpha \leqslant X < \beta]$,
এবং $P[a \leqslant X \leqslant b] = 1$.

এরূপ অপেক্ষক $f$-কে বলা হয় অবিচ্ছিন্ন সম্ভাবনা চল $X$-এর **সম্ভাবনা ঘনত্ব অপেক্ষক** (probability density function).

উভয়বিধ চলের ক্ষেত্রেই $F(x) = P[X \leqslant x]$-এর মান নির্দেশক অপেক্ষক $F$-কে বলা হয় $X$ চলের **বিভাজন অপেক্ষক** (distribution function)। অবিচ্ছিন্ন চলের ক্ষেত্রে উল্লেখ্য যে, সব $x$-এর জন্যে $f(x) = \frac{d}{dx} F(x)$.

এখন যদি $\int_{a}^{b} |x| f(x)\,dx < +\infty$ হয়, তাহলে

$$\mu = \int_{a}^{b} x f(x)\,dx = \int_{a}^{b} x\,dF\text{-কে}$$

অবিচ্ছিন্ন চল $X$-এর গাণিতিক প্রত্যাশা বলা হয়। এছাড়া,

$$\sigma^2 = \int_{a}^{b} (x-\mu)^2 f(x)\,dx = \int_{a}^{b} (x-\mu)^2 dF\text{-কে}$$

বলে $X$-এর ভেদমান এবং একে $E(X-\mu)^2$ সংকেতসূত্রেও প্রকাশ করা হয়। আমরা সর্বদাই ধ'রে নেব যে, আমাদের আলোচ্য যে কোন অবিচ্ছিন্ন চল $X$-এর জন্যেই ওপরে বর্ণিত ধর্মবিশিষ্ট সম্ভাবনা ঘনত্ব অপেক্ষক $f$-এর অস্তিত্ব থাকবে। আমরা জানি যে পরিসংখ্যা $f_i$ সমন্বিত কতিপয় মান $x_i (i = 1, 2, \ldots)$-এর যৌগিক গড় হচ্ছে

$$\bar{x} = \sum x_i \frac{f_i}{n}, \quad n = \sum f_i \text{ লিখে}$$

এখন যদি একটি বিচ্ছিন্ন সম্ভাবনা চল $X$-এর জন্যে $P[X=x_i]=p_i$ $(i=1, 2, \ldots)$ হয় তবে $p_i$-কে $\frac{f_i}{n}$-এর একটি সীমামান হিসেবে গণ্য করা যায়। কাজেই $X$-এর গাণিতিক প্রত্যাশা $\mu=\sum_i x_i p_i$ থেকে $n$ অর্থাৎ পরীক্ষণের পুনরনুষ্ঠান সংখ্যা বৃদ্ধির সঙ্গে সঙ্গে $\bar{x}$-এর মান পরিণামে কী রকম দাঁড়াবে তার একটি ইঙ্গিত পাওয়া যায়। এই জন্যে গাণিতিক প্রত্যাশাকে অনেক সময় সম্ভাবনা চলের গড় (average) ব'লেও বর্ণনা করা হয়। অবিচ্ছিন্ন চলের ক্ষেত্রেও গাণিতিক প্রত্যাশাকে যৌগিক গড়ের পরিণত রূপ হিসেবে দেখা যায়।

## 7.13 গাণিতিক প্রত্যাশা-সংক্রান্ত উদাহরণমালা :

**উদা. 7.19** একটি সুষম ছক্কা উৎক্ষিপ্ত হলে তাতে যে সংখ্যানির্দেশক চিহ্ন দেখা যাবে তার গাণিতিক প্রত্যাশা কত?

ছক্কাটি সুষম। তাই এতে যে কোন সংখ্যানির্দেশক চিহ্ন দেখতে পাওয়ার সম্ভাবনা হচ্ছে $\frac{1}{6}$. এখন, মনে কর, ছক্কাটিতে 1, 2, 3, 4, 5 ও 6-সূচক চিহ্ন দৃষ্ট হওয়ার সঙ্গে সঙ্গে একটি সম্ভাবনা চল $X$-কে যথাক্রমে 1, 2, 3, 4, 5 ও 6 এই কটি মান আরোপ করা হবে।

তাহলে, $$E(X)=\frac{1}{6}\sum_{i=1}^{6} i=\frac{7}{2}$$

হচ্ছে নির্ণেয় গাণিতিক প্রত্যাশা। এই মান থেকে একটি আভাস পাওয়া যায় ছক্কাটি বহুবার নিক্ষিপ্ত হলে তাতে $X$-এর গড় মান আনুমানিক কত হবে। তেমনি একটি সুষম মুদ্রা বহুবার উৎক্ষিপ্ত হলে বলা যাবে যে তাতে যে অনুপাতে সম্মুখপার্শ্ব দেখা যাবে তার মান হচ্ছে $\frac{1}{2}$ কারণ $\frac{1}{2}$ হচ্ছে মুদ্রায় সম্মুখ-পার্শ্ব দৃষ্ট হওয়ার প্রত্যাশিত পরিসংখ্যা।

**উদা. 7.20** এক ব্যক্তিকে $A$, $B$ ও $C$ এই তিনটি বিভিন্ন জাতের সিগারেটের তিনটি মোড়ক থেকে তিনটি সিগারেট নিয়ে তাদের ধূমপান ক'রে কোন্ সিগারেটটি কোন্ জাতের তা অনুমান করতে অনুরোধ করা হলে শুদ্ধরূপে অনুমিত সিগারেটের প্রত্যাশিত সংখ্যা কত?

ধরা যাক, পরপর যে তিনজাতের সিগারেটের ধূম পরীক্ষা করা হ'ল সেগুলি যথাক্রমে $A$, $B$ ও $C$. কিন্তু ধূমপানের ভিত্তিতে ঐ ব্যক্তি তাদের জাত বলতে পারেন যথাক্রমে (1) $A, B, C$, (2) $A, C, B$, (3) $B, A, C$, (4) $B, C, A$, (5) $C, A, B$ এবং (6) $C, B, A$. বলা বাহুল্য আর কোনরকম সিদ্ধান্ত সম্ভব নয়। তাহলে এই সব ক্ষেত্রে শুদ্ধ অনুমিতির সংখ্যা যথাক্রমে 3, 1, 1, 0, 0 এবং 1. আমরা ধ'রে নেব যে, অনুমানগুলি সব সমান সম্ভাবনার সাহায্যেই করা হচ্ছে ; অর্থাৎ প্রত্যেক অনুমানের সম্ভাবনা $\frac{1}{6}$. এখন সঠিকভাবে অনুমিত সিগারেটের সংখ্যাকে একটি সম্ভাবনা চল $X$ দ্বারা নির্দেশ করা হলে $P[X=3] = \frac{1}{6}$, $P[X=2]=0$, $P[X=1]=\frac{3}{6}=\frac{1}{2}$, $P[X=0]=\frac{2}{6}=\frac{1}{3}$.

সুতরাং $E(X)=3\times\frac{1}{6}+2\times 0+1\times\frac{1}{2}+0\times\frac{1}{3}=1.$

## 7.14 দুটি সম্ভাবনাশ্রয়ী চলের যুগ্ম-বিভাজন (joint distribution of two random variables) :

মনে করা যাক, যে কোন সম্ভাবনাভিত্তিক পরীক্ষণ-সংশ্লিষ্ট কোন ঘটনা ঘটবার সঙ্গে সঙ্গে দুটি পৃথক্ প্রকৃতমানাশ্রয়ী চলকে তাদের নিজ নিজ মানসীমার মধ্যে এক একটি মান আরোপ করা হচ্ছে। উদাহরণস্বরূপ ধরা যাক, একটি ছক্কা উৎক্ষিপ্ত হলে তার ফলাফল অনুযায়ী দুটি চল $X$ ও $Y$ নিম্নোক্তরূপে নির্দিষ্ট হ'ল :

$X=i$ ($i=1, 2, \ldots, 6$) যদি ছক্কার ওপর দৃষ্ট চিহ্ন $i$-সংখ্যার নির্দেশক হয়,

$Y = \begin{cases} 0 & \text{যদি ছক্কায় দৃষ্ট চিহ্ন বিযুগ্ম-সংখ্যা নির্দেশ করে,} \\ i \ (i=2, 4, 6) & \text{যদি ছক্কায় দৃষ্ট চিহ্ন } i\text{-সংখ্যা নির্দেশ করে।} \end{cases}$

তাহলে ছক্কার প্রতিটি নিক্ষেপণে অবেক্ষিত ঘটনা অনুসারে $X$ ও $Y$ তাদের স্ব স্ব মানসীমায় এক এক জোড়া মান গ্রহণ করবে। কাজেই $X$ ও $Y$-এর ঐরূপ প্রতিজোড়া মান গ্রহণের ব্যাপারটি হচ্ছে এক একটি সম্ভাবনাত্মক ঘটনা। কাজেই উল্লিখিত $X$ ও $Y$-কে যুক্তভাবে সম্ভাবনাশ্রয়ী চল ব'লে গ্রহণ করতে পারি। এস্থলে অবশ্য $X$ নিজে এবং $Y$ নিজে পৃথক্‌ভাবে এক একটি সম্ভাবনাশ্রয়ী চল। বর্তমান উদাহরণটিতে চল-দুটি যে সমস্ত মান গ্রহণ করে, তার সম্ভাবনা নীচের সারণীতে দেখানো যেতে পারে।

### $X$ ও $Y$-এর যুগ্ম-সম্ভাবনা-বিভাজন

| $Y \backslash X$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 0 | $\frac{1}{6}$ | 0 | $\frac{1}{6}$ | 0 | $\frac{1}{6}$ | 0 |
| 2 | 0 | $\frac{1}{6}$ | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | $\frac{1}{6}$ | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | $\frac{1}{6}$ |

এখানে উদাহরণতঃ $P[X=1, Y=0]=\frac{1}{6}$, $P[X=2, Y=0]=0$,
$P[X=3, Y=0]=\frac{1}{6}$, $P[X=4, Y=4]=\frac{1}{6}$ ইত্যাদি।

সাধারণভাবে বলা যায় যে, কোন সম্ভাবনাসাপেক্ষ পরীক্ষণ-সংশ্লিষ্ট মৌলিক ঘটনা ঘটবার সঙ্গে সঙ্গে যদি দুটি বিচ্ছিন্ন চল $X$ এবং $Y$-কে তাদের মানসীমার মধ্যে যথাক্রমে $x_i$ ও $y_j$ ($i=1, 2, \ldots$ ; $j=1, 2, \ldots$) মান-দুটি একই সঙ্গে আরোপ করা হয় যাতে $X$ এবং $Y$ তাদের এরকম মান গ্রহণ করার ব্যাপারটিকে একটি সম্ভাব্য ঘটনা বলা যায়, তবে আমরা বলি যে, $X$ এবং $Y$ হচ্ছে যুক্তভাবে সম্ভাবনাশ্রয়ী বিচ্ছিন্ন চল এবং আমরা এই চলটিকে $(X, Y)$ এই দ্বৈতসত্তার সাহায্যে প্রকাশ ক'রে থাকি। যদি $X$ এবং $Y$-এর মানগুলি যথাক্রমে $x_1, \ldots x_n$ এবং $y_1, \ldots y_m$ হয় তবে সংক্ষেপে লেখা যায় $P[X=x_i, Y=y_j] =p_{ij}$, $i=1, \ldots n$, $j=1, \ldots m$ এবং আমরা 7.1 সারণীটি গঠন করতে পারি।

এই সারণীতে $p_{ij}$ ($i=1, \ldots n$ ; $j=1, \ldots m$) মানগুলি সম্মিলিতভাবে যুক্ত-সম্ভাবনাশ্রয়ী $(X, Y)$ চলটির সম্ভাবনা-বিভাজনটি নির্দেশ করে, কারণ এরাই দেখায় $X$ এবং $Y$-এর বিভিন্ন মানদ্বৈত $(x_i, y_j)$ গুলির মধ্যে পূর্ণ সম্ভাবনা

$$1=\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}=\sum_{i=1}^{n}\sum_{j=1}^{m} P[X=x_i, Y=y_j]$$

কি-ভাবে নিবেশিত রয়েছে। এবার আমরা লিখিব,

$$p_{i.}=\sum_{j=1}^{m} p_{ij}=\sum_{j=1}^{m} P[X=x_i, Y=y_j]=P[X=x_i].$$

### সারণী 7.1

### যুক্ত-সম্ভাবনাশ্রয়ী বিচ্ছিন্ন চলদ্বয় X ও Y-এর সম্ভাবনা-বিভাজন

| $X \backslash Y$ | $y_1$ | $y_2$ | $\cdots$ | $y_j$ | $\cdots$ | $y_m$ | প্রান্তীয় সমষ্টি |
|---|---|---|---|---|---|---|---|
| $x_1$ | $p_{11}$ | $p_{12}$ | $\cdots$ | $p_{1j}$ | $\cdots$ | $p_{1m}$ | $p_{1\cdot}$ |
| $x_2$ | $p_{21}$ | $p_{22}$ | $\cdots$ | $p_{2j}$ | $\cdots$ | $p_{2m}$ | $p_{2\cdot}$ |
| $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ |
| $x_i$ | $p_{i1}$ | $p_{i2}$ | $\cdots$ | $p_{ij}$ | $\cdots$ | $p_{im}$ | $p_{i\cdot}$ |
| $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ |
| $x_n$ | $p_{n1}$ | $p_{n2}$ | $\cdots$ | $p_{nj}$ | $\cdots$ | $p_{nm}$ | $p_{n\cdot}$ |
| প্রান্তীয় সমষ্টি | $p_{\cdot 1}$ | $p_{\cdot 2}$ | $\cdots$ | $p_{\cdot j}$ | $\cdots$ | $p_{\cdot m}$ | 1 |

কারণ, $[Y=y_1]$, $[Y=y_2]$, ..., $[Y=y_m]$ ঘটনাগুলি পরস্পরব্যতিরেকী ও পরস্পরনিঃশেষী, যার ফলে লেখা যায়,

$$P[X=x_i,\ Y=y_1, y_2, \ldots, y_m\text{-এর মধ্যে যে কোন একটি}\ ]=P[X=x_i].$$

ঠিক তেমনিভাবে, $p_{\cdot j}=\sum_{i=1}^{n} p_{ij}=\sum_{i=1}^{n} P[X=x_i,\ \ Y=y_j]=P[Y=y_j].$

তাহলে, স্পষ্টতঃই $p_{i\cdot}$ মানগুলি কেবলমাত্র $X$ চলের সম্ভাবনা-বিভাজন সূচিত করে, কারণ এরা দেখায় পূর্ণ সম্ভাবনা 1 কি-ভাবে কেবলমাত্র $X$-এর মানগুলির মধ্যে রয়েছে, $Y$-এর মানগুলি যাই হোক না কেন। পরিভাষানুযায়ী বলা হয় যে, $p_{i\cdot}$ মানগুলি $X$-এর প্রান্তীয় বিভাজন (marginal distribution) নির্দেশ করে। অনুরূপভাবে, $p_{\cdot j}$ মানগুলি কেবলমাত্র $Y$ চলের সম্ভাবনা-বিভাজন সূচিত করে এবং আমরা বলি যে, তারা $Y$-এর প্রান্তীয় বিভাজন নির্দেশ করে। ওপরের সারণীতে $(X, Y)$-এর যুগ্ম বিভাজন এবং $X$ ও $Y$-এর প্রান্তীয় বিভাজন ছাড়া আরও দুই শ্রেণীর সম্ভাবনা-বিভাজন প্রচ্ছন্ন রয়েছে। যে কোন একটি লব্ধ-

পঙক্তি (column) ( ধরা যাক $j$-তম ) নেওয়া যাক। তাতে $p_{1j}$, $p_{2j}$, ...$p_{ij}$, ..., $p_{nj}$ এই মানগুলি রয়েছে। এদের সমষ্টি হচ্ছে $p_{.j}$.

এখন, $\frac{p_{1j}}{p_{.j}}, \frac{p_{2j}}{p_{.j}}, \ldots, \frac{p_{ij}}{p_{.j}}, \ldots, \frac{p_{nj}}{p_{.j}}$ ($p_{.j} > 0$ ধ'রে )

এই অনুপাতগুলির দিকে মন দেওয়া যাক। এদের সমষ্টি হচ্ছে 1 এবং এরা দেখায় পূর্ণ সম্ভাবনা 1 কি-ভাবে $X$-এর বিভিন্ন মান $x_1, \ldots x_i, \ldots, x_n$-এর মধ্যে নিবেশিত রয়েছে যদি এটা মেনে নেওয়া হয় যে, $Y$ তার বিভিন্ন মানের মধ্যে কেবলমাত্র একটি অর্থাৎ $y_j$-কে আশ্রয় ক'রে আছে।

$$\text{এখন, } \frac{p_{ij}}{p_{.j}} = \frac{P[X = x_i, Y = y_j]}{P[Y = y_j]}$$

সংখ্যাটি $Y$ চলটি তার একটি মাত্র মান $y_j$-কে আশ্রয় ক'রে আছে এই সর্তাধীনে $X$ চল তার $x_i$ মান ধারণ করার সর্তাধীন সম্ভাবনা নির্দেশ করে। তাই বলা যায় যে, বিভিন্ন $i = 1, \ldots n$-এর জন্যে $\frac{p_{ij}}{p_{.j}}$ রাশিগুলি একযোগে $Y$-এর মান $y_j$-তে স্থির রয়েছে এই সর্তাধীনে $X$-এর সর্তাধীন-সম্ভাবনা-বিভাজন (conditional probability distribution) সূচিত করে। প্রত্যেক $j = 1, \ldots, m$-এর জন্যে এরকম এক একটি সর্তাধীন সম্ভাবনা-বিভাজন আছে। এদেরকে পঙক্তি-বিভাজনও (array distribution) বলা হয়। ঠিক এমনিভাবে আমরা বলতে পারি $i$-তম শায়ী পঙক্তিতে (row array) যে মানগুলি $p_{i1}, p_{i2}, \ldots p_{ij}$, ...$p_{im}$ রয়েছে তাদের সবাইকে তাদের সমষ্টি $p_{i.}$ (ধনাত্মক ধ'রে) দিয়ে ভাগ ক'রে যে মানগুলি $\frac{p_{i1}}{p_{i.}}, \frac{p_{i2}}{p_{i.}}, \ldots, \frac{p_{ij}}{p_{i.}}, \ldots, \frac{p_{im}}{p_{i.}}$ পাওয়া যায়, তারা $X$-এর মান $x_i$-তে স্থির রয়েছে এই সর্তাধীনে $Y$ চলের সর্তাধীন সম্ভাবনা-বিভাজন নির্দেশ করে। প্রত্যেক $i = 1, \ldots, n$-এর জন্যে এমনি এক একটি শায়ী পঙক্তি-বিভাজন রয়েছে। এখানে আরও উল্লেখ্য যে,

$$F(x, y) = \sum_{x_i \leqslant x} \sum_{y_j \leqslant y} p_{ij} = \sum_{x_i \leqslant x} \sum_{y_j \leqslant y} P[X = x_i, Y = y_j]$$

লিখলে, $F$-কে বলা হবে যুক্ত সম্ভাবনা দ্বৈতচল $(X, Y)$-এর বিভাজন অপেক্ষক। তেমনি, $F_1(x) = \sum_{x_i \leqslant x} p_{i.}$ ও $F_2(y) = \sum_{y_j \leqslant y} p_{.j}$ লিখলে $F_1$ ও $F_2$-কে

যথাক্রমে $X$ ও $Y$-এর প্রান্তীয়-বিভাজন-অপেক্ষক (marginal distribution fuction) বলে। এছাড়া,

$$G_1(x|j) = \sum_{x_i \leqslant x} \frac{p_{ij}}{p_{\cdot j}} \text{ ও } H_2(y|i) = \sum_{y_j \leqslant y} \frac{p_{ij}}{p_{i\cdot}} \text{ লিখলে}$$

$G_1$ ও $H_2$-কে যথাক্রমে $Y=y_j$ এই সর্তাধীনে $X$-এর এবং $X=x_i$ এই সর্তাধীনে $Y$-এর সর্তাধীন-বিভাজন-অপেক্ষক (conditional distribution functions) বলা হয়।

ওপরে যে $n$ ও $m$ এর উল্লেখ করা হয়েছে তারা সসীমসংখ্যা নাও হতে পারে। কিন্তু তবু (অর্থাৎ তাদের একটি বা উভয়ে অসীমাভিসারী হলেও) ওপরের সংজ্ঞাগুলির গঠনে কোন পরিবর্তন হয় না; একমাত্র ব্যতিক্রম এই যে, তখন $i=1, 2, \ldots n, \ldots$ ইত্যাদি এবং $j=1, 2, \ldots, m, \ldots$ ইত্যাদি লেখা হবে।

এখন, মনে কর, $X$ ও $Y$ উভয়েই অবিচ্ছিন্ন চল এবং তাদের মান যথাক্রমে $[\alpha, \beta]$ ও $[\gamma, \delta]$ অন্তরের মধ্যে সীমাবদ্ধ। এই অন্তরদ্বয় অবশ্য $(-\infty, +\infty)$-এর সমান হতে আপত্তি নেই। এক্ষেত্রে $P[\alpha \leqslant X \leqslant \beta]=1$, $P[\gamma \leqslant Y \leqslant \delta] =1$ ও $P[\alpha \leqslant X \leqslant \beta, \gamma \leqslant Y \leqslant \delta]=1$. এখন যদি কোন সম্ভাবনাসাপেক্ষ পরীক্ষণ-এর ভিত্তিতে অবেক্ষিত মৌলিক ঘটনাবলীর বিভিন্নতা অনুযায়ী $X$ ও $Y$-কে যুগপৎ যথাক্রমে $[\alpha, \beta]$-এর কোন উপ-অন্তর $[a, b]$ ও $[\gamma, \delta]$-এর কোন উপ-অন্তর $[c, d]$-এর মধ্যে কোন মান আরোপ করা হয় তাহলে আমরা বলব যে, $(X, Y)$ একটি অবিচ্ছিন্ন সম্ভাবনাশ্রয়ী দ্বৈতচল। অবশ্য এক্ষেত্রে আমরা স্বীকার করে নেব যে একটি অবিচ্ছিন্ন দ্বিচলবিশিষ্ট অপেক্ষক $f$-এর অস্তিত্ব রয়েছে যার জন্যে নিম্নলিখিত (i) ও (ii) এই সর্ত-দুটি সর্বদা খাটে। এই $f$ কে $(X, Y)$ দ্বৈত বা যুগল চলের সম্ভাবনা-ঘনত্ব-অপেক্ষক (probability density function) বলা হবে। সর্ত-দুটি হ'ল :

(i) প্রত্যেক $x, y$-এর জন্যে $f(x, y) \geqslant 0$

এবং (ii) $\int_\alpha^\beta \int_\gamma^\delta f(x, y)\, dy\, dx = 1.$

এক্ষেত্রে আমরা আরও ধ'রে নেব যে,

$P[X \leqslant x, Y \leqslant y] = \int_\alpha^x \int_\gamma^y f(x, y) dy\, dx = F(x, y)$ লিখলে প্রত্যেক

$x$ ও $y$ এর জন্যে $\frac{\delta^2 F(x, y)}{\delta x \delta y}$-এর অস্তিত্ব রয়েছে এবং $f(x, y) = \frac{\delta^2 F(x, y)}{\delta x \delta y}$. এই $F$-কে $(X, Y)$ এর যুগ্ম-বিভাজন-অপেক্ষক বলা হবে। এখানে উল্লেখযোগ্য যে, $\int_a^b \int_c^d f(x, y)\, dy\, dx = \int_a^b \int_c^d dF(x, y) = P[a < X < b, c < Y < d]$ হচ্ছে $X$ এবং $Y$ চল-দুটির মান যুগপৎ যথাক্রমে $[a, b]$ ও $[c, d]$ অন্তর-দুটির মধ্যে থাকার সম্ভাবনা। এছাড়া,

$$g(x) = \int_\gamma^\delta f(x, y)\, dy \text{ ও } h(y) = \int_\alpha^\beta f(x, y)\, dx$$ লিখলে

যথাক্রমে $g$ ও $h$-কে $X$ এবং $Y$-এর প্রান্তীয় সম্ভাবনা-বিভাজনের সম্ভাবনা-ঘনত্ব-অপেক্ষক বলা হবে। আবার,

$$F_1(x) = \int_\alpha^x g(t)\, dt \text{ ও } F_2(y) = \int_\gamma^y h(u)\, du$$ লিখলে $F_1$ ও $F_2$-কে যথাক্রমে $X$ এবং $Y$-এর প্রান্তীয়-বিভাজন-অপেক্ষক বলা হয়। তাছাড়া $g(x) > 0$ ও $h(y) > 0$ হলে,

$$f_1(y|x) = \frac{f(x, y)}{g(x)} \text{ ও } f_2(x|y) = \frac{f(x, y)}{h(y)}$$ লিখে $f_1$ ও $f_2$-কে যথাক্রমে $X$-এর $x$ মানে নির্ণীত $Y$-এর সর্তাধীন সম্ভাবনা-ঘনত্ব-অপেক্ষক (conditional probability density function) এবং $Y$-এর $y$ মানে নির্ণীত $X$-এর সর্তাধীন সম্ভাবনা-ঘনত্ব-অপেক্ষক বলা হয়। সবশেষে $F_1(y|x) = \int_\gamma^y f_1(u|x)\, du$ ও $F_2(x|y) = \int_\alpha^x f_2(t|y)\, dt$ লিখে $F_1$-কে $X$-এর $x$ মানে নির্ণীত $Y$-এর সর্তাধীন-বিভাজন-অপেক্ষক (conditional distribution function) এবং $F_2$ কে $Y$ এর $y$ মানে নির্ণীত $X$ এর সর্তাধীন-বিভাজন-অপেক্ষক বলা হয়।

## 7.15 সম্ভাবনাশ্রয়ী চলের স্বাতন্ত্র্য বা অনধীনতা (stochastic independence of random variables)

ধরা যাক, $X$ এবং $Y$ দুটি বিচ্ছিন্ন সম্ভাবনাশ্রয়ী চল এবং এদের সম্ভাবনা-বিভাজন যথাক্রমে $p_{io} = P[X = x_i]$, $i = 1, 2, \ldots n$, এবং $p_{oj} = P[Y = y_j]$, $j = 1, 2, \ldots m, \ldots$ দ্বারা প্রকাশ করা হোক। এখানে অবশ্যই

$\sum_{i=1} p_{io} = 1$ এবং $\sum_{j=1} p_{oj} = 1$. আরও ধরা যাক যে, এদের যুগ্ম-সম্ভাবনা-বিভাজন $p_{ij} = P[X = x_i,\ Y = y_j]$, [ এখানে $i = 1,\ 2, \ldots n, \ldots$ ও $j = 1,\ 2, \ldots m, \ldots$ এবং $\sum_i \sum_j p_{ij} = 1$] দ্বারা প্রকাশিত। এখন, $[X = x_i]$ ঘটনাকে $A_i$ ও $[Y = y_j]$ ঘটনাকে $B_j$ দ্বারা নির্দেশ করলে মিশ্র ঘটনা $[X = x_i,\ Y = y_j]$ কে $A_i \cap B_j$ দ্বারা সূচিত করা যায়। আমরা জানি যে,

$$P(A_i \cap B_j) = P(A_i).P(B_j) \qquad \cdots \qquad \cdots \quad (7.49)$$

অর্থাৎ $p_{ij} = p_{io} \times p_{oj}$

হলে $A_i$ ও $B_j$ ঘটনাদ্বয় পরস্পর সম্ভাবনাগত অর্থে অনধীন। দুটি ঘটনার স্বাতন্ত্র্য বা অনধীনতার এই সংজ্ঞাটিকেই একটু প্রসারিত ক'রে বলা হয় যে, যদি

প্রত্যেক $i,\ j = 1,\ 2, \cdots \infty$-এর জন্যে $p_{ij} = p_{io} \times p_{oj} \qquad \cdots \quad (7.50)$

হয়, তাহলে $X$ এবং $Y$ সম্ভাবনা চল-দুটি পরস্পর অনধীন। এখন যদি $U(X)$ ও $V(Y)$ যথাক্রমে $X$ এবং $Y$-এর যে কোন অপেক্ষক হয় এবং তাদের যদি একটি যুগ্ম-সম্ভাবনা-বিভাজন থাকে যা $X$ ও $Y$-এর যুগ্ম-সম্ভাবনা-বিভাজনের মাধ্যমে সুনির্দিষ্ট, তাহলে $X$ ও $Y$ পরস্পর অনধীন হলে $U(X)$ এবং $V(Y)$ও পরস্পর অনধীন হবে।

এখন ধরা যাক যে, $X$ ও $Y$ দুটি অবিচ্ছিন্ন চল এবং $g$ ও $h$ যথাক্রমে তাদের সম্ভাবনা-ঘনত্ব-অপেক্ষক। তাহলে, $[a < X < b]$ ও $[c < Y < d]$ ঘটনা-দুটির সম্ভাবনা হচ্ছে যথাক্রমে

$$P[a < X < b] = \int_a^b g(x)\, dx$$

ও $$P[c < Y < d] = \int_c^d h(y)\, dy.$$

এখন $(X,\ Y)$ হচ্ছে একটি অবিচ্ছিন্ন দ্বৈতচল এবং $f$ তার সম্ভাবনা-ঘনত্ব-অপেক্ষক হলে মিশ্র ঘটনা $[a < X < b,\ c < Y < d]$-এর সম্ভাবনা হচ্ছে

$$P[a < X < b,\ c < Y < d] = \int_a^b \int_c^d f(x,\ y)\, dy\, dx.$$

যদি $F$, $F_1$ ও $F_2$ যথাক্রমে $(X, Y)$, $X$ ও $Y$-এর যুগ্ম ও প্রান্তীয় বিভাজন-অপেক্ষক হয়, এবং যদি প্রত্যেক $x$, $y$-এর জন্যে

$$f(x, y) = g(x)\, h(y) \qquad \cdots \quad (7.51)$$

অথবা $$F(x, y) = F_1(x)\, F_2(y) \qquad \cdots \quad (7.52)$$

হয়, তাহলে $X$ ও $Y$-কে পরস্পর অনধীন বলা হবে।

তিন বা ততোধিক সম্ভাবনা চলের যুগ্ম-বিভাজন, প্রান্তীয় ও সর্তাধীন বিভাজন ইত্যাদি ও তাদের যৌথভাবে পরস্পর অনধীনতাও ওপরে বর্ণিতভাবে অগ্রসর হয়ে নির্দিষ্ট করা যায়। সম্ভাবনা চলের বিভাজন, ঘটনার সম্ভাবনার দ্বারা নির্দিষ্ট একথা স্মরণে রেখে তিন বা ততোধিক ঘটনার পরস্পর অনধীনতা যেমনভাবে নির্দিষ্ট হয়েছিল, তিন বা ততোধিক সম্ভাবনা চলের অনধীনতার সংজ্ঞাও তেমনি ভাবে নির্দিষ্ট করতে হবে। বাস্তবিক, এটা দেখানো যাবে যে, কতগুলি সম্ভাবনা চল পরস্পর যুগ্মভাবে অনধীন হলেও যৌথভাবে তারা পরস্পর অনধীন নাও হতে পারে।

## 7.16 গাণিতিক প্রত্যাশার যৌগিক সূত্র (sum law of mathematical expectation) :

**নির্বচন :** ধর, দুটি সম্ভাবনা চল $X$ ও $Y$ এর গাণিতিক প্রত্যাশা যথাক্রমে $E(X)$ ও $E(Y)$. তাহলে, সম্ভাবনা চল $(X+Y)$-এর গাণিতিক প্রত্যাশা $E(X+Y)$ হবে $E(X)+E(Y)$-এর সমান। অর্থাৎ দুটি সম্ভাবনা চলের সমষ্টির গাণিতিক প্রত্যাশা হচ্ছে তাদের একক গাণিতিক প্রত্যাশাদ্বয়ের সমষ্টি।

**প্রমাণ :**

প্রথম ক্ষেত্র : উভয় চলই বিচ্ছিন্ন।

ধরা যাক, $X$-এর মানগুলি $x_1, x_2, \ldots, x_n, \ldots$ ও $p_1, p_2, \ldots, p_n, \ldots$ যথাক্রমে তাদের সম্ভাবনা এবং $Y$-এর মানগুলি হচ্ছে $y_1, y_2, \ldots, y_m, \ldots$ ও $q_1, q_2, \ldots, q_m, \ldots$ যথাক্রমে তাদের সম্ভাবনা।

অর্থাৎ $P[X=x_i] = p_i \; (i=1, 2, \ldots,$ ও $P[Y=y_j] = q_j \; (j=1, 2, \ldots).$

তাহলে $E(X) = \sum_i x_i p_i$ ও $E(Y) = \sum_j y_j q_j$.

আবার মনে করা যাক, $p_{ij} = P[X=x_i, \; Y=y_j]$। তাহলে $(X+Y)$

চলটি $(x_i+y_j)$ মান গ্রহণ করার ঘটনা $[X=x_i,\ Y=y_j]$ ঘটনার সঙ্গে অভিন্ন। তাই $P[X+Y=x_i+y_j]=P[X=x_i,\ Y=y_j]=p_{ij}$ অর্থাৎ $(X+Y)$-এর সম্ভাবনা-বিভাজন $p_{ij}$ মানগুলির মাধ্যমে নির্ধারিত। কাজেই, সংজ্ঞানুযায়ী, $(X+Y)$-এর গাণিতিক প্রত্যাশা হচ্ছে

$$E(X+Y)=\sum_i\sum_j (x_i+y_j)\,p_{ij}$$

$$=\sum_i\sum_j x_ip_{ij}+\sum_i\sum_j y_j\,p_{ij}$$

$$=\sum_i x_i\sum_j p_{ij}+\sum_j y_j\sum_i p_{ij}.$$

এখন, $$\sum p_{ij}=P[X=x_i,\ Y=y_1]+P[X=x_i,\ Y=y_2]+\cdots$$

$$=P[X=x_i]=p_i,$$

এবং $$\sum p_{ij}=P[X=x_1,\ Y=y_j]+P[X=x_2,\ Y=y_j]+$$

$$=P[Y=y_j]=q_j.$$

সুতরাং $$E(X+Y)=\sum_i x_ip_i+\sum_j y_jq_j=E(X)+E(Y)$$

অর্থাৎ $$E(X+Y)=E(X)+E(Y). \qquad \cdots \quad (7.53)$$

**দ্বিতীয় ক্ষেত্র :** উভয় চলই অবিচ্ছিন্ন এবং সম্ভাবনা-ঘনত্ব-যুক্ত।

ধরা যাক, $X$ ও $Y$ দুটি অবিচ্ছিন্ন সম্ভাবনা চল এবং $g$ ও $h$ যথাক্রমে তাদের সম্ভাবনা-ঘনত্ব-অপেক্ষক। আরও ধরা যাক যে, $f$ তাদের যুগ্ম-সম্ভাবনা-বিভাজনের ঘনত্ব-অপেক্ষক এবং $(\alpha,\beta)$ ও $(\gamma,\delta)$ যথাক্রমে তাদের মানসীমা অর্থাৎ $P[\alpha\leqslant X\leqslant\beta]=1$ ও $P[\gamma\leqslant Y\leqslant\delta]=1$ তাহলে, $(X+Y)$ একটি অবিচ্ছিন্ন সম্ভাবনা চল এবং এর বিভাজন $(X,Y)$ দ্বৈতচলের বিভাজনের মাধ্যমে স্থিরীকৃত। ফলে,

$$E(X+Y)=\int_\alpha^\beta\int_\gamma^\delta (x+y)f(x,y)\,dy\,dx$$

$$=\int_\alpha^\beta\int_\gamma^\delta xf(x,y)dy\,dx+\int_\alpha^\beta\int_\gamma^\delta yf(x,y)\,dy\,dx$$

$$=\int_{\alpha}^{\beta} x\left[\int_{\gamma}^{\delta} f(x,y)dy\right]dx+\int_{\gamma}^{\delta} y\left[\int_{\alpha}^{\beta} f(x,y)\,dx\right]dy$$

$$=\int_{\alpha}^{\beta} x\,g(x)\,dx+\int_{\gamma}^{\delta} yh(y)\,dy=E(X)+E(Y)$$

অর্থাৎ $E(X+Y)=E(X)+E(Y)$. ··· (7.54)

এই সূত্রটিকে আরও সম্প্রসারিত করা যায়। দুটির পরিবর্তে যে কোন সসীম সংখ্যক চলের ক্ষেত্রেই এই সূত্রটি সত্য। অর্থাৎ $X_1,\ldots, X_k$ যদি $k$ সংখ্যক সম্ভাবনা চল এবং তাদের গাণিতিক প্রত্যাশা যথাক্রমে

$E(X_1),\ldots\ldots, E(X_k)$ হয়, তাহলে

$$E(X_1+\cdots+X_k)=E(X_1)+\cdots+E(X_k). \quad \cdots \quad (7.55)$$

**প্রমাণ :** $k$-এর মান 3 হলে

$E(X_1+X_2+X_3)=E[(X_1+X_2)+X_3]=E(X_1+X_2)+E(X_3)$

[ পূর্ববর্তী সূত্র (7.53) ও (7.54) অনুসারে ],

$=E(X_1)+E(X_2)+E(X_3)$ [ পূর্ববর্তী সূত্র (7.53) ও (7.54) অনুসারে ]।

এরপর আরোহ পদ্ধতি প্রয়োগ কর।

**টীকা :** (7.53), (7.54) ও (7.55)-কে গাণিতিক প্রত্যাশার যৌগিক সূত্র বলা হয়।

## 7.17 গাণিতিক প্রত্যাশার গুণন সূত্র (product law of mathematical expectation) :

**নির্বচন :** দুটি পরস্পর অনধীন সম্ভাবনা চল $X$ ও $Y$-এর গাণিতিক প্রত্যাশা $E(X)$ ও $E(Y)$ হলে $XY$-এর গাণিতিক প্রত্যাশা $E(XY)$ হবে $E(X)\,E(Y)$-এর সমান।

**প্রমাণ :**

প্রথম ক্ষেত্র : উভয় চলই বিচ্ছিন্ন।

ধরা যাক, $X$ চলের মানগুলি $x_1, x_2, \ldots, x_n, \ldots$ ও তাদের সম্ভাবনা যথাক্রমে $p_1, p_2, \ldots p_n, \ldots$ এবং $Y$-এর মানগুলি $y_1, y_2, \ldots, y_m, \ldots$ ও তাদের সম্ভাবনা যথাক্রমে $q_1, q_2, \ldots\ldots q_m, \ldots$। আরও ধরা যাক,

$$p_{ij}=P[X=x_i,\ Y=y_j].$$

তাহলে, $(X, Y)$ দ্বৈতচলটির সম্ভাবনা-বিভাজন $p_{ij}$ মানগুলির মাধ্যমে

নির্ধারিত। এখন, $XY$ একটি সম্ভাবনা চল এবং $[XY=x_i\ y_j]$ ও $[X=x_i, Y=y_j]$ ঘটনা-তুটি সমতুল অর্থাৎ সমসম্ভব অর্থাৎ $P[XY=x_i y_j]$ $=P[X=x_i, X=y_j]=p_{ij}$ এখন যেহেতু $X$ ও $Y$ পরস্পর অনধীন, কাজেই প্রত্যেক $i, j=1, 2,\ldots$এর জন্যেই

$$P[X=x_i, Y=y_j]=P[X=x_i]P[Y=y_j]$$

অর্থাৎ $p_{ij}=p_i\ q_j$.

তাহলে, সংজ্ঞানুযায়ী :

$$E(XY)=\sum_i \sum_j x_i\ y_j\ P[XY=x_i\ y_j]$$

$$=\sum_i \sum_j x_i\ y_j\ p_{ij}=\sum_i \sum_j x_i\ y_j\ p_i\ q_j$$

$$=\sum_i x_i\ p_i \sum_j y_j\ q_j=\left(\sum_i x_i\ p_i\right)\left(\sum_f q_j\ y_j\right)$$

$$=E(X)\ E(Y)$$

অর্থাৎ $E(XY)=E(X)\ E(Y)$. ... (7.56)

দ্বিতীয় ক্ষেত্র : উভয় চলই অবিচ্ছিন্ন এবং সম্ভাবনা-ঘনত্ব-যুক্ত।

ধর, তুটি পরস্পর অনধীন সম্ভাবনা চল $X$ ও $Y$-এর সম্ভাবনা-ঘনত্ব-অপেক্ষক যথাক্রমে $g$ ও $h$ এবং তাদের যুগ্মবিভাজনের ঘনত্ব-অপেক্ষক হচ্ছে $f$. তাহলে প্রত্যেক $x, y$-এর জন্যে $f(x, y)=g(x)\ n(y)$. আরও ধরা যাক $[\alpha, \beta]$ ও $[\gamma, \delta]$ যথাক্রমে $X$ ও $Y$-এর মানসীমা অর্থাৎ

$$P[\alpha \leqslant X \leqslant \beta]=1 \text{ ও } P[\gamma \leqslant Y \leqslant \delta]=1.$$

এখন $XY$ চলের বিভাজন $(X, Y)$ দ্বৈত চলের বিভাজনের মাধ্যমে স্থিরীকৃত। তাহলে, সংজ্ঞানুযায়ী,

$$E(XY)=\int_\alpha^\beta \int_\gamma^\delta xy\ f\ (x, y)\ dy\ dx=\int_\alpha^\beta \int_\gamma^\delta xy\ g(x)\ h(y)\ dy\ dx$$

$$=\int_\alpha^\beta xg(x)\left[\int_\gamma^\delta y\ h(y)\ dy\right]dx=\int_\alpha^\beta x\ g(x)\ E(Y)dx$$

$$=E(Y)\int_\alpha^\beta x\ g(x)\ dx=E(Y).E(X),$$

অর্থাৎ $E(XY)=E(X)\ E(Y)$. ... (7.57)

এই সূত্রটিকে আরও সম্প্রসারিত করা যায়। দুটির পরিবর্তে যে কোন (সসীম) $k$ সংখ্যক সম্ভাবনা চল $X_1, \ldots, X_k$ যদি পরস্পর যৌথভাবে অনধীন হয় এবং $E(X_1), \ldots, E(X_k)$ যথাক্রমে তাদের গাণিতিক প্রত্যাশা হয়, তাহলে সম্ভাবনা চল $X_1, \ldots, X_k$-এর গাণিতিক প্রত্যাশা

$$E(X_1 \ldots X_k) = E(X_1) \ldots E(X_k) \qquad \cdots \quad (7.58)$$

**প্রমাণ :** $k$-এর মান 3 হলে

$$E(X_1X_2X_3) = E[(X_1X_2)X_3] = E(X_1X_2)\ E(X_3)$$
$$= E(X_1)\ E(X_2)\ E(X_3). \quad [(7.56) \text{ ও } (7.57) \text{ দ্রষ্টব্য }]$$

এর পর আরোহ-পদ্ধতি প্রযোজ্য

(7.56), (7.57) ও (7.58)-কে গাণিতিক প্রত্যাশার গুণন সূত্র বলা হয়।

**টীকা (1) :** ধ্রুবকের গাণিতিক প্রত্যাশা।

ধরা যাক, $X$ এমন একটি সম্ভাবনা চল যার সম্ভাবনা-বিভাজন এরূপ যে,

$$P[X=c]=1 \text{ ও } P[X \neq c]=0.$$

তাহলে সংজ্ঞানুযায়ী, $E(X) = c \times P[X=c] = c.1 = c.$

প্রচলিত রীতি (convention) অনুযায়ী কোন ধ্রুবক $c$-এর গাণিতিক প্রত্যাশা বলতে $E(c) = E(X) = c$-কেই ধরা হয়। এখানে $X$ হচ্ছে উল্লিখিত-রূপে নিবেশিত চল।

**টীকা (2) :** যদি $c$ একটি ধ্রুবক এবং $X$ যে কোন সম্ভাবনা চল হয়, তাহলে $cX$ একটি সম্ভাবনা চল এবং এর সম্ভাবনা-বিভাজন $X$-এর সম্ভাবনা-বিভাজন দ্বারা এককভাবে নির্দিষ্ট (uniquely determined) এবং ধরা হয় যে এর গাণিতিক প্রত্যাশা হচ্ছে

$$E(cX) = cE(X). \qquad \cdots \quad (7.59)$$

**টীকা (3) :** $X$ যে কোন সম্ভাবনা চল এবং $a$ ও $b$ যে কোন দুটি ধ্রুবক হলে সংজ্ঞানুযায়ী $E(aX+b) = aE(X)+b. \qquad \cdots \quad (7.60)$

## 7.18 সহভেদমান ও ভেদমান (covariance and variance) :

**সংজ্ঞা :** যে কোন দুটি সম্ভাবনা চল $X$ ও $Y$-এর গাণিতিক প্রত্যাশা $E(X)$ ও $E(Y)$ হলে $[X-E(X)\ Y-E(Y)]$ একটি সম্ভাবনা চল এবং এর গাণিতিক

প্রত্যাশা হচ্ছে $E[X-E(X)\ Y-E(Y)]$. একে $X$ ও $Y$-এর সহভেদমান (covariance) বলা হয় এবং cov $(X, Y)$ সংকেতসূত্রে প্রকাশ করা হয়।

**উপপাদ্য 5.** দুটি সম্ভাবনা চল $X$ ও $Y$ পরস্পর অনধীন হলে তাদের সহভেদমানের মান হবে শূন্য।

**প্রমাণ :** সংজ্ঞানুযায়ী,

$$\text{cov}\,(X, Y)=E[X-E(X)\ Y-E(Y)]$$
$$=E[XY-XE(Y)-YE(X)+E(X)\,E(Y)]$$
$$=E(XY)-E(X)\,E(Y)-E(Y)E(X)+E(X)E(Y)$$

[$E(X)$ ও $E(Y)$ হচ্ছে ধ্রুবক ]

$$=E(XY)-E(X)\,E(Y)=0$$ [(7.56) ও (7.57) স্মর্তব্য ]।

**উপপাদ্য 6.** যদি সম্ভাবনা চল $X_i$ $(i=1,\ldots,k)$-এর ভেদমান $V(X_i)$ এবং $X_i$ ও $X_j$ $(i\neq j=1, 2,\ldots)$ এর সহভেদমান cov $(X_i, X_j)$ হয়, তাহলে

$$V\left(\sum_{i=1}^{k} X_i\right)=\sum_{i=1}^{k} V(X_i)+\sum_{\substack{i=1\\ i\neq j}}^{k}\sum_{j=1}^{k}\text{cov}\,(X_i, X_j) \qquad \cdots \quad (7.61)$$

হবে।

**প্রমাণ :** $$V\left(\sum_{i=1}^{k} X_i\right)=E\left[\sum_{i=1}^{k} X_i-E\left(\sum_{i=1}^{k}X_i\right)\right]^2$$

$$=E\left[\sum_{i=1}^{k}\left(X_i-E(X_i)\right)\right]^2 \qquad (7.55 \text{ দ্রষ্টব্য})$$

$$=\sum_{i=1}^{k} E[X_i-E(X_i)]^2+\sum_{\substack{i=1\\ i\neq j}}^{k}\sum_{j=1}^{k} E[(X_i-E(X_i))(X_j-E(X_j))]$$

$$=\sum_{i=1}^{k} V(X_i)+\sum_{\substack{i=1\\ i\neq j}}^{k}\sum_{j=1}^{k}\text{cov}\,(X_i, X_j). \qquad \cdots \quad (7.62)$$

**অনুসিদ্ধান্ত :** যদি প্রত্যেক $X_i$ ও $X_j$ $(i\neq j=1,\ldots k)$ সম্ভাবনাতত্ত্বানুযায়ী পরস্পর অনধীন হয়, তবে $V\left(\sum_{i=1}^{k} X_i\right)=\sum_{i=1}^{k} V(X_i)$.

**প্রমাণ :** উপপাদ্য 3 ব্যবহার ক'রে পাওয়া যায়

প্রত্যেক $i \neq j = 1, \ldots, k$-এর জন্যে $\text{cov}(X_i, X_j) = 0$,

এখন (7.62) থেকে স্পষ্টতঃই পাওয়া যায় $V\left(\sum_{i=1}^{k} X_i\right) = \sum_{i=1}^{k} V(X_i)$.

## 7.19 চেবিশেফের সহায়ক উপপাদ্য (Chebyshev's lemma) :

**নির্বচন :** একটি অ-ঋণাত্মক মানাশ্রয়ী (non-negative valued) সম্ভাবনা চল $\mathsf{U}$-এর গাণিতিক প্রত্যাশা $\theta$ হলে যে কোন প্রকৃত মানাশ্রয়ী সংখ্যা $t$ ($\neq 0$)-এর জন্যে

$$P\left[\mathsf{U} > \theta t^2\right] < \frac{1}{t^2} \qquad \cdots \quad (7.63)$$

হবে।

**প্রমাণ :** $\theta = 0$ হলে উপপাদ্যটি প্রমাণের অপেক্ষা রাখে না। কারণ, সেক্ষেত্রে $P[\mathsf{U} = 0] = 1$ এবং $P[\mathsf{U} > 0] = 0 < \frac{1}{t^2}$, প্রত্যেক প্রকৃত মানাশ্রয়ী $t$ ($\neq 0$)-এর জন্যে। কাজেই আমরা $\theta > 0$ ধ'রে নেব। তাছাড়া আমরা $\mathsf{U}$-কে একটি বিচ্ছিন্ন চল ধ'রে প্রতিজ্ঞাটি প্রমাণ করব, যদিও আসলে দেখানো যায় যে, এটি সাধারণভাবেও সত্য।

ধরা যাক, $\mathsf{U}$-এর মানগুলি এমনভাবে নির্দেশক সংখ্যা দ্বারা সূচিত হ'ল যে তারা দাঁড়াল $u_1, \ldots u_k, u_{k+1}, \ldots u_n, \ldots$ এবং কোন বিশেষ $t$ ($\neq 0$) এর জন্যে

$$u_i > \theta t^2 \text{ যখন } i = 1, 2, \ldots k \qquad \cdots \quad (7.64)$$

এবং
$$u_i \leqslant \theta t^2 \text{ যখন } i = k+1, k+2, \ldots, n \qquad \cdots \quad (7.65)$$

এখন, $P[\mathsf{U} = u_i] = p_i$ লিখলে পাওয়া যায়

$$\theta = E(\mathsf{U}) = \sum_{i=1}^{\infty} u_i p_i = \sum_{i=1}^{k} u_i p_i + \sum_{i=k+1}^{\infty} u_i p_i$$

$$> \sum_{i=1}^{k} u_i p_i \quad [\text{ যেহেতু প্রত্যেক } i = 1, 2, \cdots \text{ এর জন্যে } u_i > 0 \text{ ও } p_i > 0]$$

$$\theta t^2 \sum_{i=1}^{k} p_i \qquad [(7.64) \text{ স্মর্তব্য }]$$

$$= \theta t^2 P[\text{U} = u_1 \text{ অথবা } u_2, \ldots \text{ অথবা } u_k]$$

$$= \theta t^2 P[\text{U} > \theta t^2]$$

সুতরাং $P[\text{U} > \theta t^2] < \frac{1}{t^2}$.

(7.63) কে

$$P[\text{U} \leqslant \theta t^2] > 1 - \frac{1}{t^2} \qquad (7.66)$$

আকারেও লেখা যায়, কারণ $P[\text{U} > \theta t^2] + P[\text{U} \leqslant \theta t^2] = 1$

এবং $P[\text{U} \leqslant \theta t^2] = 1 - P[\text{U} > \theta t^2] > 1 - \frac{1}{t^2}$, [(7.63) দ্রষ্টব্য ]।

**অনুসিদ্ধান্ত :** (7.63)-তে যদি $\theta \neq 0$ ধ'রে নেওয়া হয়, তবে চেবিশেফের সহায়ক উপপাদ্যকে বিকল্পে $P[\text{U} \geqslant \theta t^2] \leqslant 1/t^2$ ... (7.67)
এই আকারেও লেখা যায় [ এর প্রমাণ নিজে দেওয়ার চেষ্টা কর ]।

## 7.20 চেবিশেফের অসমতা সম্পর্ক (Chebyshev's inequality) :

**নির্বচন :** একটি সম্ভাবনা চল $X$-এর গাণিতিক প্রত্যাশা ও ভেদমান যথাক্রমে $\mu$ ও $\sigma^2$ হলে প্রত্যেক ধনাত্মক $t$-এর জন্যে

$$P[\mu - \sigma t \leqslant X \leqslant \mu + \sigma t] > \; - \frac{1}{t^2}. \qquad \ldots \quad (7.68)$$

**প্রমাণ :** $X$-কে বিচ্ছিন্ন চল ব'লে ধ'রে নেওয়া হবে যদিও প্রতিজ্ঞাটি সাধারণভাবেও সিদ্ধ।

এখন $\text{U} = (X - \mu)^2$ লিখলে U একটি অ-ঋণাত্মক সম্ভাবনা চল এব $E(\text{U}) = E(X - \mu)^2 = \text{U}(X) = \sigma^2$.

তাহলে চেবিশেফের সহায়ক উপপাদ্য থেকে পাই

$P[(X - \mu)^2 > \sigma^2 t^2] < \frac{1}{t^2}$, প্রত্যেক $t(\neq 0)$-এর জন্যে

অর্থাৎ $P[(X-\mu)^2 \leqslant \sigma^2 t^2] > 1-\frac{1}{t^2}$

অর্থাৎ $P[|X-\mu| \leqslant \sigma t] > 1-\frac{1}{t^2}$, $t > 0$ ধরলে,

অর্থাৎ $P[\mu-\sigma t \leqslant X \leqslant \mu+\sigma t] > 1-\frac{1}{t^2}$.

মন্তব্য : (7.67) স্মরণে রেখে বলা যায় যে, যেহেতু $\sigma \neq 0$, আমরা লিখতে পারি $P[(X-\mu)^2 \geqslant \sigma^2 t^2] \leqslant \frac{1}{t^2}$. ··· (7.69)

## 7.21 বৃহৎ সংখ্যা-বিধি (law of large numbers) :

ধরা যাক, $\{X_n\}$ হচ্ছে যে কোন পরীক্ষণ সংশ্লিষ্ট সম্ভাবনা চল $X_n$-এর একটি পরম্পরা (sequence) এবং $c$ যে কোন একটি ধ্রুবক। তাহলে যে কোন ধনাত্মক রাশি $\in$-এর জন্যে পরীক্ষণ $\varepsilon$-এর সঙ্গে সংশ্লিষ্ট যে সমস্ত মৌলিক ঘটনা $\omega$-এর জন্যে $|X_n(\omega)-c| > \in$ হবে তাদের গুচ্ছ অর্থাৎ $A=\{\omega : |X_n(\omega)-c| > \in\}$ একটি ঘটনা নির্দেশ করবে। সংক্ষেপে $A$-কে আমরা $A=[|X_n-c|> \in]$ এই সংকেতসূত্র সাহায্যেও প্রকাশ করব। এখন যদি $n$ ক্রমাগত বাড়তে থাকে এবং সেই সঙ্গে $P(A)=P[|X_n-c|> \varepsilon]$-এর পরিমাণ ক্রমাগত ছোট হতে হতে শূন্যের কাছাকাছি এগিয়ে যায়, তাহলে আমরা বলব যে $\{X_n\}$ পরম্পরাটি সম্ভাবনাতত্ত্বানুযায়ী বা সম্ভাবনাত্মকভাবে ধ্রুবক $c$-এর অভিসারী হয় এবং আমরা লিখি

$$X_n \xrightarrow{P} c.$$

এক্ষেত্রে বলা যাবে যে, $n$ যদি একটি নির্দিষ্ট সংখ্যা $n_0=n_0(\in, \delta)$-এর চেয়ে বড় হয়, তবে $\delta(>0)$ যত ছোটই হোক,

$$P[|X_n-c|> \in] < \delta.$$

এখানে $n_0=n_0(\in, \delta)$ সংখ্যাটি $\varepsilon$ এবং $\delta$-এর ওপর নির্ভর করতে পারে।

এখানে উল্লেখযোগ্য যে, যদি এমন হয় যে প্রত্যেক $\omega$-এর জন্যে $n > n_0(\in, \delta)$ হলেই $|X_n(\omega)-c| \leqslant \in$, তবে বলা হবে যে, $\{X_n\}$ পরম্পরাটি $c$ ধ্রুবকের প্রতি গাণিতিক বা গাণিতিক বিশ্লেষণ সূত্রানুযায়ী অগ্রসর হয় (converges analytically to $c$). কিন্তু $\{X_n\}$ পরম্পরাটি সম্ভাবনাত্মকভাবে

$c$-এর প্রতি অগ্রসর হওয়ার অর্থ এই যে, $n$ এর মান $n_0(\varepsilon, \delta)$ অপেক্ষা বড় হওয়া সত্ত্বেও কোন কোন মৌলিক ঘটনা $\omega$-এর জন্যে $|X_n(\omega) - c|$-এর মান $\varepsilon$-এর চেয়ে ছোট নাও হতে পারে ; কিন্তু যে সমস্ত মৌলিক ঘটনা $\omega$-এর জন্যে এ ব্যাপারটি ঘটবে তাদের সমবায়ে গঠিত ঘটনার সম্ভাবনা $\delta$-এর চেয়ে ছোট হবে।

এখন, ধরা যাক $\{X_n\}$ একটি সম্ভাবনাসাপেক্ষ চলের পরম্পরা এবং $\{\mu_n\}$ অপর একটি পরম্পরা যার প্রত্যেকটি উপাদানই এক একটি ধ্রুবক। এখন, একটি নতুন পরম্পরা $\{D_n\}$ এমন ভাবে গঠন করা যাক যে,

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i,\ \xi_n = \frac{1}{n}\sum_{i=1}^{n} \mu_i \text{ এবং } D_n = S_n - \xi_n.$$

এখন যদি $D_n \xrightarrow{P} 0$ হয়, অর্থাৎ $\{D_n\}$ পরম্পরাটি যদি সম্ভাবনাত্মকভাবে 0-এর অভিসারী হয়, তবে আমরা বলব যে, $\{X_n\}$ পরম্পরাটি **সামান্য বৃহৎ সংখ্যা-বিধি** (weak law of large numbers) মেনে চলে।

উল্লেখ্য যে, $\{\mu_n\}$ পরম্পরাটির প্রতিটি উপাদান একটি মাত্র সংখ্যা $\mu$-এর সমান হতে পারে। সেক্ষেত্রে $D_n = \frac{1}{n}\sum_{i=1}^{n} X_i - \mu$ হবে। বৃহৎ সংখ্যা-বিধির সংজ্ঞায় এছাড়া আর কোন পরিবর্তন হবে না।

## 7.22 বৃহৎ সংখ্যা-বিধির প্রয়োগ :

**উদা. 7.21**

ধরা যাক $\{X_n\}$ একটি সম্ভাবনাসাপেক্ষ চলের পরম্পরা। এখানে সম্ভাবনা চলগুলির ধর্ম এমন যে, প্রত্যেক $k$-এর জন্যে $E(X_k) = \mu_k$-এর অস্তিত্ব রয়েছে।

লেখা যাক $S_n = \sum_{i=1}^{n} X_i$ ও $V(S_n) = B_n$.

তাহলে,

$$\lim_{n\to\infty} \frac{B_n}{n^2} = 0 \tag{7.70}$$

হলে $\{X_n\}$ পরম্পরাটি সামান্য বৃহৎ সংখ্যা-বিধি মেনে চলে।

**প্রমাণ ঃ** লেখা যাক

$$D_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_i) = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mu_i.$$

তাহলে, আমাদের দেখাতে হবে যে, $D_n \xrightarrow{P} 0.$

দেখা যাচ্ছে যে, $E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu_i$

$$V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\,V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\,V(S_n) = \frac{B_n}{n^2}.$$

তাহলে, চেবিশেফের অসমতা সম্পর্ক ব্যবহার ক'রে পাই

$$P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mu_i\right| > \epsilon\right] < \frac{B_n}{n^2\epsilon}$$

অর্থাৎ $P[|D_n| > \epsilon] < \frac{B_n}{n^2\epsilon^2}.$

সুতরাং $\lim_{n\to\infty} P[|D_n| > \epsilon] < \lim_{n\to\infty} \frac{B_n}{n^2\epsilon^2} = \frac{1}{\epsilon^2}\lim_{n\to\infty}\frac{B_n}{n^2} = 0$

অর্থাৎ $\lim_{n\to\infty} P[|D_n| > \epsilon] = 0.$

কাজেই $D_n \xrightarrow{P} 0$

অর্থাৎ (7.70) এ উল্লিখিত সর্তাধীনে $\{X_n\}$ পরম্পরাটি সামান্য বৃহৎ সংখ্যাবিধি মেনে চলে। এই ব্যাপারটিকে একটি উপপাদ্য ব'লে ধরা যায়। একে চেবিশেফের বৃহৎ-সংখ্যা বিধি বলা হয়।

**উদা. 7.22** ধরা যাক $\{X_n\}$ হচ্ছে এমন একটি পরস্পর অনধীন সম্ভাবনা-চলের পরম্পরা যে, প্রত্যেক $i = 1, 2, \ldots, n, \ldots$-এর জন্যে

$$P[X_i = 1] = 1 \text{ ও } P[X_i = 0] = 1 - p = q.$$

এখন, $D_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - p)$ লিখলে দেখা যাবে যে,

$D_n \xrightarrow{P} 0$ অর্থাৎ $\{X_n\}$ পরম্পরাটি সামান্য বৃহৎ-সংখ্যা বিধি মেনে চলে।

**প্রমাণ :** $E(X_i)=p,\ V(X_i)=p(1-p).$

কাজেই $E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)=p,\quad V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)=\frac{p(1-p)}{n}$

এবং চেবিশেফের অসমতা সম্পর্ক থেকে পাই [ (7.69) দ্রষ্টব্য ]

$$P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i-p\right|>\epsilon\right]<\frac{p(1-p)}{n\epsilon^2}=\frac{pq}{n\epsilon^2}<\frac{1}{4n\epsilon^2}$$

সুতরাং $\lim_{n\to\infty} P[|D_n|>\epsilon]<\frac{1}{4\epsilon^2}\lim_{n\to\infty}\frac{1}{n}=0$

অর্থাৎ, $D_n \xrightarrow{P} 0.$

এখানে, $pq \leq \frac{1}{4}$ কারণ $(p+q)^2-(p-q)^2=4pq$

অর্থাৎ, $4pq=1-(p-q)^2$ অর্থাৎ $4pq$-এর গরিষ্ঠ মান হচ্ছে 1.

## 7.23 পৌনঃপুনিক প্রয়াস (repeated trials) ও বেরনুলির উপপাদ্য :

যদি কার্যতঃ অভিন্ন পরিস্থিতিতে একটি সম্ভাবনাসাপেক্ষ পরীক্ষণ বারবার অনুষ্ঠিত হতে থাকে যাতে প্রত্যেক অনুষ্ঠানে আনুসঙ্গিক সম্ভাব্য ঘটনাগুলির প্রকৃতি অভিন্ন থাকে, তাহলে ঐ পরীক্ষণগুলিকে একত্রে পৌনঃপুনিক প্রয়াস বা প্রচেষ্টা বলা হয়। পৌনঃপুনিক প্রয়াসে পরীক্ষণের যে কোন অনুষ্ঠান-সংশ্লিষ্ট কোন ঘটনা যদি ঐ পরীক্ষণের অপর কোন অনুষ্ঠানে ঘটিত যে কোন ঘটনার ( এমন কি ঘটনা-দুটি অভিন্ন হলেও ) অনধীন হয় তবে প্রচেষ্টাগুলিকে অনধীন (independent repeated trials) বলা হয়। যদি পরস্পর অনধীন পৌনঃপুনিক প্রচেষ্টাগুলির প্রকৃতি এমন হয় যে, প্রত্যেক পরীক্ষণ বা প্রয়াসে লক্ষণীয় মৌলিক ঘটনা থাকে মাত্র দুটি, যার একটিকে পরিভাষানুযায়ী 'সাফল্য' (success) এবং অপরটিকে ব্যর্থতা (failure) বলা হয়, এবং প্রত্যেক প্রয়াসে ( বা পরীক্ষণের অনুষ্ঠানে ) 'সাফল্য' ( ব্যর্থতা ) লাভের সম্ভাবনা সর্বদা কোন সংখ্যা $p\ (q=1-p)$-এর সমান থাকে, এই পৌনঃপুনিক প্রচেষ্টারাজিকে **বেরণুলির প্রয়াস** (Bernoullian trials) বলে। যেমন, একটি ছক্কা যদি বারবার নিক্ষিপ্ত হয় এবং তাতে যে কোন সময় 6 সূচক চিহ্নযুক্ত প্রান্ত ওপরে থাকাকে সাফল্য ও অপর কোন প্রান্ত ওপরে থাকাকে যদি ব্যর্থতা বলা হয়, তাহলে একটি বেরণুলীয় প্রকৃতির পৌনঃপুনিক প্রয়াসের সারি গঠিত হ'ল বলা যায়।

**বেরণুলির উপপাদ্য :**

**নির্বচন :** কোন সম্ভাবনাসাপেক্ষ পরীক্ষণ ε যদি বারবার অন্ততঃ কার্যতঃ সদৃশ সর্তাধীনে নিষ্পন্ন হয় এবং এই পরীক্ষণের অনুষ্ঠান পরম্পরার প্রথম $n$টিতে অবেক্ষিত কোন ঘটনা $E$-এর পরিসংখ্যা যদি $f_n$ হয়, তবে $n$-এর মান বৃদ্ধির সঙ্গে সঙ্গে $P\left[\left|\frac{f_n}{n}-p\right|>\eta\right]$-এর মান ক্রমশঃ শূন্যাভিসারী হয়। এখানে $p$ হচ্ছে পরীক্ষণটির যে কোন অনুষ্ঠানে $E$ ঘটনার সম্ভাবনা এবং $\eta$ যে কোন একটি ধনাত্মক সংখ্যা।

**প্রমাণ :** উদাহরণ 7.22-এর সাহায্য নিয়ে নিজে চেষ্টা কর।

**টীকা :** বেরণুলির উপপাদ্য থেকে কোন ঘটনার অবেক্ষিত পরিসংখ্যা-অনুপাতকে সেই ঘটনার সম্ভাবনার প্রাক্কলক হিসেবে কেন গণ্য করা হয় তার একটি যুক্তি পাওয়া যায়। পরিসংখ্যা অনুপাতটি ঘটনার সম্ভাবনার প্রতি সম্ভাবনাগত অর্থে ক্রমাভিসারী হয়।

## 7.24 বিবিধ উদাহরণমালা :

**উদা. 7.23** ছটি বিভিন্ন পুরো তাসের প্যাকেটের প্রত্যেকটি থেকে একটি ক'রে তাস সমসম্ভব উপায়ে বেছে নিলে তিনটি লাল ও তিনটি কালো তাস পাওয়ার সম্ভাবনা কত ?

প্রতিটি প্যাকেট থেকে এক একটি তাস নিয়ে সেটির রঙ সাদা কি কালো পরীক্ষা ক'রে দেখাকে যদি পরীক্ষণ প্রচেষ্টা বলা হয় তাহলে এখানে ছটি বেরণুলীয় প্রয়াস গঠিত হচ্ছে। নির্বাচিত তাসের রঙ লাল (কালো) হলে প্রয়াসটি 'সার্থক' হ'ল বলা যেতে পারে। প্রতিটি প্যাকেটে 52টি তাসের মধ্যে 26টি লাল ও 26টি কালো তাস থাকে। কাজেই বলতে পারি যে প্রত্যেক প্রচেষ্টায় 'সার্থকতা' লাভের সম্ভাবনা $=\frac{26}{52}=\frac{1}{2}$ এবং 'ব্যর্থতার' সম্ভাবনাও অবশ্যই $\frac{1}{2}$. তাহলে ছটি প্রচেষ্টায় তিনটি সার্থকতা লাভের সম্ভাবনা নির্ণয় করতে হবে। উল্লিখিত ঘটনাটি $\binom{6}{3}$ সংখ্যক বিভিন্ন উপায়ে ঘটতে পারে কারণ 6টি প্রচেষ্টার যে 3টিতে সার্থকতা ঘটবে তাদের $\binom{6}{3}$ সংখ্যক উপায়ে বেছে নেওয়া যাবে। আবার যে প্রতিটি ক্ষেত্রে 3টি সার্থকতা ও 3টি ব্যর্থতা ঘটে তাকে একটি ঘটনা বললে তার সম্ভাবনা হচ্ছে $(\frac{1}{2})^3(\frac{1}{2})^3$ কারণ 6টি প্রচেষ্টার প্রত্যেকটিতে সার্থকতা বা ব্যর্থতার ঘটনা পরস্পর অনধীন। সুতরাং নির্ণেয় সম্ভাবনা হচ্ছে $\binom{6}{3}(\frac{1}{2})^3(\frac{1}{2})^3=\frac{5}{16}$.

**উদা. 7.24** 1, 2, ..., 6 সংখ্যাচিহ্নিত একটি সুষম ছক্কা চারবার নিক্ষিপ্ত হলে লব্ধ মোট সংখ্যা 14 হওয়ার সম্ভাবনা কত ?

এখানে প্রতিটি ছক্কার জন্যে মোট সমসম্ভব পরিস্থিতিসংখ্যা 6. কাজেই চারটি ছক্কায় লক্ষণীয় মোট পরিস্থিতিসংখ্যা $6^4$.

প্রশ্নে উল্লিখিত ঘটনাটির অনুকূল পরিস্থিতিসংখ্যা নির্ণয় করতে গিয়ে আমরা লক্ষ্য করতে পারি যে, যদি $i_1, \ldots, i_4$ যথাক্রমে প্রথম, দ্বিতীয়, তৃতীয় ও চতুর্থ ছক্কায় দৃষ্ট সংখ্যা নির্দেশ করে, তাহলে আমাদের নির্ণয় করতে হবে

$$i_1+i_2+i_3+i_4=14$$

এই সর্তসাপেক্ষে $i_1, i_2, i_3, i_4$-এর প্রত্যেকে কতরকম বিভিন্ন উপায়ে 1, 2, ..., 6—এই মানগুলি গ্রহণ করতে পারে।

এখন, $(x^1+x^2+x^3+x^4+x^5+x^6)^4$-এর প্রসারণটির কথা বিবেচনা করলে দেখা যায় যে, এর বিভিন্ন পদগুলি হ'ল $x^{i_1+i_2+i_3+i_4}$ ; এতে $i_1, i_2, i_3$ ও $i_4$-এর প্রত্যেকেই হচ্ছে 1, 2, ..., 6. এছাড়া $(x^1+\cdots+x^6)^4$-এ $x^{14}$-এর সহগ হচ্ছে ঠিক যতরকমভাবে $i_1+i_2+i_3+i_4=14$ সর্তসাপেক্ষে $i_1, i_2, i_3$ ও $i_4$-কে 1, 2, 3, 4, 5, 6 এই মানগুলি আরোপ করা যায়। তাহলে, নির্ণেয় পরিস্থিতিসংখ্যা হচ্ছে

$(x+x^2+\cdots+x^6)^4$-এর প্রসারণে $x^{14}$-এর সহগ

অর্থাৎ $x^4(1+x+\cdots+x^5)^4$ " " $x^{14}$ "

" $(1+x+\cdots+x^5)^4$ " " $x^{10}$ "

" $\left(\dfrac{1-x^6}{1-x}\right)^4$ " " $x^{10}$ "

" $(1-x^6)^4(1-x)^{-4}$ " " $x^{10}$ "

কাজেই উল্লিখিত সহগের মান হচ্ছে

$$\frac{4\times5\times6\times7\times8\times9\times10\times11\times12\times13}{1\times2\times3\times4\times5\times6\times7\times8\times9\times10}$$

$$-4\times\frac{4\times5\times6\times7}{1\times2\times3\times4}=286-140=146.$$

সুতরাং নির্ণেয় সম্ভাবনা হ'ল $P=\dfrac{146}{6^4}=\dfrac{73}{648}$.

**উদা. 7.25** একটি খেলায় জিতবার জন্যে $A$ এবং $B$ দুটি ক'রে ছক্কা নিক্ষেপ করতে থাকে যতক্ষণ না তাদের মধ্যে একজন প্রথমে একটি ছয়

এবং একটি এক পায়। $A$-র দুটি ছক্কাই সুষম; কিন্তু $B$-এর একটি ছক্কা সুষম, কিন্তু অন্যটির এক প্রান্ত এমনভাবে ভারযুক্ত যে তাতে ছয় অন্য যে কোন সংখ্যার চেয়ে তিনগুণ বেশী অনুপাতে দৃষ্ট হয়। $A$-র জয়ের সম্ভাবনা কত হবে (1) যদি সে সুরু করে এবং (2) যদি $B$ সুরু করে?

বলা বাহুল্য যে, কোন একবার দুটি ছক্কা নিক্ষেপ করলে $A$ তাতে 6 ও 1 পাবে এমন সম্ভাবনা হচ্ছে $\frac{1}{6}\times\frac{1}{6}+\frac{1}{6}\times\frac{1}{6}=\frac{1}{18}$. পক্ষান্তরে, $B$ যে তার নিক্ষিপ্ত ছক্কা-দুটিতে 6 ও 1 পাবে তার সম্ভাবনা হচ্ছে $\frac{3}{8}\times\frac{1}{6}+\frac{1}{8}\times\frac{1}{6}=\frac{1}{12}$, কারণ যেহেতু অসম ছক্কাটিতে 6 অন্য সংখ্যার চেয়ে 3 গুণ বেশী সংখ্যায় দৃষ্ট হয় আমরা ধ'রে নেব যে প্রতি 8 বার নিক্ষেপণে গড়ে 6 দৃষ্ট হবে 3 বার ও 1, 2, ..., 5 দৃষ্ট হবে 1 বার ক'রে অর্থাৎ অসম ছক্কাটিতে 6 পাবার সম্ভাবনা $\frac{3}{8}$ এবং অন্যান্য সংখ্যা 1, 2, ..., 5-এর প্রত্যেকটি দৃষ্ট হওয়ার সম্ভাবনা $\frac{1}{8}$. এখন, প্রথম ক্ষেত্রে অর্থাৎ যখন $A$ সুরু করে, তখন $A$ জয়ী হতে পারে নিম্নোক্ত পরস্পর ব্যতিরেকী বিকল্প পরিস্থিতিগুলিতে, যথা: (i) প্রথমেই $A$-র সাফল্য, (ii) $A$ ব্যর্থ, তারপর $B$ ব্যর্থ ও তারপর $A$-র সাফল্য, (iii) প্রথম দু'বার $A$ ও $B$ উভয়েই ব্যর্থ ও তৃতীয় প্রচেষ্টায় $A$ সফল, (iv) ইত্যাদি।

তাহলে প্রতিবার দুটি ছক্কা ছুঁড়ে ছয় ও এক পাবার প্রচেষ্টাগুলি পরস্পর অনধীন একথা মনে রেখে এবং সম্ভাবনার যৌগিক ও গুণন সূত্র প্রয়োগ ক'রে পাওয়া যাবে

$A$ সুরু করলে তার জয়ী হবার সম্ভাবনা $P_1$ হচ্ছে

$$P_1=\tfrac{1}{18}+\tfrac{17}{18}\cdot\tfrac{11}{12}\cdot\tfrac{1}{18}+\left(\tfrac{17}{18}\cdot\tfrac{11}{12}\right)^2\cdot\tfrac{1}{18}+\left(\tfrac{17}{18}\cdot\tfrac{11}{12}\right)^3\tfrac{1}{18}+\cdots$$

$$=\tfrac{1}{18}\left[1+\left(\tfrac{17}{18}\cdot\tfrac{11}{12}\right)+\left(\tfrac{17}{18}\cdot\tfrac{11}{12}\right)^2+\cdots\right]=\tfrac{1}{18}\cdot\frac{1}{1-\frac{17}{18}\cdot\frac{11}{12}}=\tfrac{12}{29}.$$

অনুরূপে, দ্বিতীয় ক্ষেত্রে অর্থাৎ $B$ সুরু করলে $A$-এর জয়ী হবার সম্ভাবনা $P_2$-এর মান পাওয়া যাবে $P_2=\frac{11}{29}$.

**উদা. 7.26** বেরনুলীয় প্রকৃতির একটি পৌনঃপুনিক প্রয়াস পরম্পরায় সর্বপ্রথম লব্ধ সার্থকতার ঠিক পূর্বে পর্যন্ত অবেক্ষিত ব্যর্থতার প্রত্যাশিত মান কত?

ধরা যাক, $X$ হচ্ছে উল্লিখিত ব্যর্থতার সংখ্যা। তাহলে $X$ একটি সম্ভাবনা চল এবং $X$-এর মান যে কোন $x$-এর সমান হওয়ার অর্থ হচ্ছে এই যে প্রথম

$x$ সংখ্যক প্রচেষ্টায় ক্রমাগত ব্যর্থতা আসার পর $(x+1)$-তম প্রচেষ্টায় সর্বপ্রথম সাফল্যলাভ ঘটে। এখন উল্লিখিত প্রয়াস পরম্পরায় প্রথম $x$ সংখ্যক ব্যর্থতার পর $(x+1)$-তম প্রচেষ্টায় সার্থকতা লাভের ঘটনার সম্ভাবনা স্পষ্টতঃই হচ্ছে $q^x p$, অর্থাৎ $P[X=x]=q^x p$, এখানে $x$-এর সম্ভাব্যমান হচ্ছে 0, 1, 2...ইত্যাদি। সুতরাং নির্ণেয় প্রত্যাশিত সংখ্যা হচ্ছে

$$E(X)=\sum_{x=0}^{\infty} x\, q^x p = p(q+2q^2+3q^3+\cdots)$$

$$=pq\,(1+2q+3q^2+\cdots)=pq\,(1-q)^{-2}=\frac{pq}{p^2}=\frac{q}{p}.$$

**উদা. 7.27** একটি বাক্সে $k$ প্রকার সমসংখ্যক বস্তু রয়েছে। এর থেকে যদি প্রতিবার একটি ক'রে বস্তু সমসম্ভব উপায়ে বেছে নিয়ে পরবর্তী নির্বাচনের আগে সেটি ফেরৎ দেওয়া হয় এবং $n$ যদি প্রত্যেক প্রকার বস্তু অন্ততঃ একবার ক'রে গৃহীত হওয়ার জন্যে প্রয়োজনীয় ন্যূনতম নির্বাচনসংখ্যা হয়, তবে $E(n)$ ও $V(n)$-এর মান কত হবে ?

ধরা যাক, $n_i$ হচ্ছে ন্যূনতম প্রয়োজনীয় নির্বাচনসংখ্যা যাতে বিভিন্ন $(i-1)$ প্রকার বস্তু সংগৃহীত হওয়ার পর অন্য নূতনতর বস্তুর উদ্ভব ঘটে। তাহলে,

$$n=n_1+n_2+\cdots n_k.$$

সুতরাং $$E(n)=E\left(\sum_{i=1}^{k} n_i\right)=\sum_{i=1}^{k} E(n_i).$$

এখন, যেহেতু প্রতিবার নির্বাচনের পরবর্তী নির্বাচনের পূর্বে সংগৃহীত বস্তুটি আধারে পুনঃপ্রত্যর্পিত হয়, কাজেই এই নির্বাচনপ্রচেষ্টাগুলিকে পরস্পর অনধীন মনে করতে আপত্তি নেই। সুতরাং ধরা যায় যে, প্রত্যেক $i=1,$ $2,\ldots$-এর জন্যে $n_i$ গুলি পরস্পর অনধীন। সুতরাং $V\left(\sum_{i=1}^{k} n_i\right)=\sum_{i=1}^{k} V(n_i).$

এখন, প্রত্যেক $i=1,\cdots k$-এর জন্যে $n_i$ হচ্ছে 1, 2, 3, 4,$\cdots$ইত্যাদি মান ধারণক্ষম একটি সম্ভাবনা চল। $n_i$-এর মান 1, 2, 3,... হবে যদি $(i-1)$ প্রকার বস্তু সংগৃহীত হবার পর নূতনতর বস্তু সংগ্রহে যথাক্রমে 1টি, 2টি, 3টি, ... নির্বাচনের প্রয়োজন হয়। ইতিপূর্বে মোট $(i-1)$ প্রকার বস্তু নির্বাচিত হবার পর যে

কোনও নির্বাচনে নূতনতর বস্তুর নির্বাচিত হওয়ার সম্ভাবনা হচ্ছে $p_i = \frac{k-i+1}{k}$ এবং তাতে পুরাতন $(i-1)$ প্রকার বস্তুর কোনটি গৃহীত হবার সম্ভাবনা হচ্ছে

$$\frac{i-1}{k} = 1 - p_i = q_i.$$

কাজেই $P(n_i=1)=p_i,\ P(n_i=2)=q_ip_i,\ P(n_i=3)=q_i^2p_i,$

সুতরাং $E(n_i)=p_i+2q_ip_i+3q_i^2p_i+\cdots=p_i(1-q_i)^{-2}=\frac{1}{p_i}=\frac{k}{k-i+1}$

তাই $\quad E(n)=\sum_{i=1}^{k}E(n_i)=k\sum_{i=1}^{k}\frac{1}{(k-i+1)}.$

আবার, $\quad V(n_i)=E(n_i^2)-\{E(n_i)\}^2.$

কিন্তু $\quad E(n_i^2)=E[n_i(n_i-1)+n_i]=E[n_i(n_i-1)]+E(n_i)$

ফলে, $\quad V(n_i)=E[n_i(n_i-1)]+E(n_i)-\{E(n_i)\}^2.$

এখন, $E[n_i(n_i-1)]=\sum_{x=1}^{\infty}x(x-1)\,q_i^{x-1}p_i$ [ উদাহরণ 7.26 স্মর্তব্য ]

$$=2q_ip_i+3\times 2q_i^2p_i+4.3\ q_i^3p_i+\cdots$$

$$=2q_ip_i(1+3q_i+6q_i^2+\cdots)=2q_ip_i(1-q_i)^{-3}=\frac{2q_i}{p_i^2}.$$

তাই $\quad V(n_i)=\frac{2q_i}{p_i^2}+\frac{1}{p_i}-\frac{1}{p_i^2}=\frac{2q_i+p_i-1}{p_i^2}=\frac{q_i}{p_i^2}=\frac{1}{p_i^2}-\frac{1}{p_i}$

এবং $\quad V(n)=\sum_{i=1}^{k}V(n_i)=\sum_{i=1}^{k}\frac{1}{p_i^2}-\sum_{i=1}^{k}\frac{1}{p_i}$

$$=k^2\sum_{i=1}^{k}\frac{1}{(k-i+1)^2}-k\sum_{i=1}^{k}\frac{1}{(k-i+1)}$$

$$=k^2\left[1+\frac{1}{2^2}+\frac{1}{3^2}+\cdots+\frac{1}{k^2}\right]-k\left[1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{k}\right].$$

**উদা. 7.28** সাদা ও কালো রঙের কিন্তু অন্য দিক থেকে অভিন্ন বল-সমেত তিনটি বাক্স আছে ; তাদের মধ্যে সাদা ও কালো বলের সংখ্যা যথাক্রমে $a_1,\ a_2,\ a_3$ এবং $b_1,\ b_2,\ b_3$. যদি প্রত্যেকটি বাক্স থেকে সমান সম্ভাবনা

সাহায্যে একটি ক'রে বল নেওয়া হয়, তবে তাদের সব-কটি একই বর্ণের হওয়ার সম্ভাবনা কত ?

প্রতিটি বাক্স থেকে এক-একটি বল বেছে তোলা হচ্ছে এক-একটি পরীক্ষণ প্রচেষ্টা। প্রচেষ্টাগুলিকে পরস্পর অনধীন ব'লে স্বীকার ক'রে নিতে কোন আপত্তি নেই। প্রশ্নে উল্লিখিত ঘটনাটি দুটি পরস্পর ব্যতিরেকী উপায়ে ঘটতে পারে ; কারণ, বলগুলি হয় সব সাদা অথবা সব কালো হতে পারে। এখন, তিনটি বলই সাদা হওয়ার সম্ভাবনা $\frac{a_1}{a_1+b_1} \times \frac{a_2}{a_2+b_2} \times \frac{a_3}{a_3+b_3}$ এবং তিনটি বলই কালো হওয়ার সম্ভাবনা $\frac{b_1}{a_1+b_1} \times \frac{b_2}{a_2+b_2} \times \frac{b_3}{a_3+b_3}$.

সুতরাং নির্ণেয় সম্ভাবনা হচ্ছে এ দুটি সংখ্যার সমষ্টি অর্থাৎ

$$\frac{a_1\, a_2\, a_3 + b_1\, b_2\, b_3}{(a_1+b_1)(a_2+b_2)(a_3+b_3)}.$$

**উদা. 7.29** $n$-সংখ্যক ঠিকানাযুক্ত খামের প্রত্যেকটিতে এক-একটি হিসাবে $n$-সংখ্যক চিঠি যদি এলোপাথাড়িভাবে ভরার চেষ্টা করা হয়, তাহলে যথাযথ ঠিকানাযুক্ত খামে যতগুলি চিঠি ভরা হবে তাদের সংখ্যার প্রত্যাশিত মান ও ভেদমান কত হবে ?

মনে কর, খাম ও চিঠিগুলিকে পৃথক করার জন্যে তাদেরকে $1, 2, \ldots, i, \ldots, n$ সংখ্যা দিয়ে চিহ্নিত করা হ'ল। $i$-চিহ্নিত চিঠি $i$-তম খামে রাখলে চিঠিটি যথাযথ খামে রাখা হ'ল ব'লে স্বীকার করা যেতে পারে।

ধরা যাক $X_i$ $(i=1, \ldots, n)$ একটি সম্ভাবনা চল যার জন্যে

$$X_i = \begin{cases} 1, & \text{যদি } i\text{-চিহ্নিত চিঠি } i\text{-চিহ্নিত খামে রাখা হয় ;} \\ 0, & \text{অন্যথায় ;} \end{cases}$$

সুতরাং, $X = \sum_{i=1}^{n} X_i$ লিখলে $E(X)$ ও $V(X)$-এর মান নির্ণয় করতে হবে।

এখন, $P[X_i=1] = \frac{(n-1)!}{n!}$, কারণ $n$টি চিঠিকে $n$টি খামে মোট $n!$ প্রকার উপায়ে রাখা যায় এবং $i$-চিহ্নিত চিঠি $i$-চিহ্নিত খামে রাখলে বাকী $(n-1)$টি চিঠি $(n-1)$টি খামে মোট $(n-1)!$ প্রকার উপায়ে রাখা যায়।

কাজেই $P[X_i=0] = 1 - P[X_i=1] = 1 - \frac{(n-1)!}{n} \cdot 1 - \frac{1}{n}$.

আবার লক্ষণীয় যে, প্রত্যেক $i=1, \ldots, n$-এর জন্যে $X_i$ সমনিবেশিত।

এখন, $E(X_i)=1\times P[X_i=1]+0\times P[X_i=0]$

$$=1\times\frac{1}{n}+0\times\left(1-\frac{1}{n}\right)=\frac{1}{n}.$$

তাই $E(X)=E\left(\sum_1^n X_i\right)=\sum_1^n E(X_i)=n\times\frac{1}{n}=1.$

এছাড়া, $V(X_i)=E[X_i-E(X_i)]^2$

$$=\left(1-\frac{1}{n}\right)^2\times\frac{1}{n}+\left(0-\frac{1}{n}\right)^2\times\left(1-\frac{1}{n}\right)$$

$$=\frac{1}{n}\times\left(1-\frac{1}{n}\right)\left(1-\frac{1}{n}+\frac{1}{n}\right)=\frac{1}{n}\left(1-\frac{1}{n}\right).$$

এবং $\operatorname{cov}(X_i, X_j)=E[\{X_i-E(X_i)\}\{X_j-E(X_j)\}]$

$$=E(X_i X_j)-E(X_i)E(X_j)=E(X_i X_j)-\frac{1}{n^2}.$$

কিন্তু $E(X_i X_j)=1\times P[X_i=1, X_j=1]$
$+0(P[X_i=1, X_j=0]+P[X_i=0, X_j=1]$
$+P[X_i=0, X_j=0])$

$$=P[X_i=1, X_j=1]=\frac{(n-2)!}{n!}=\frac{1}{n(n-1)},$$

কারণ, $[X_i=1, X_j=1]$ ঘটনাটি ঘটবে যদি $i$- ও $j$-চিহ্নিত চিঠি-দুটি যথাক্রমে $i$- ও $j$-চিহ্নিত খামে ভরা হয় এবং বাকী $(n-2)$টি চিঠি বাকী $(n-2)$টি খামে যথেচ্ছভাবে মোট $(n-2)!$ প্রকার উপায়ে ভরা হয়।

কাজেই $\operatorname{cov}(X_i, X_j)=\frac{1}{n(n-1)}-\frac{1}{n^2}=\frac{1}{n^2(n-1)}$, $i\neq j=1, 2,\ldots n.$

সুতরাং $V(X)=\sum V(X_i)+\sum_i\sum_{j,\, i\neq j}\operatorname{cov}(X_i, X_j)$

$$=nV(X_i)+2\times\frac{n(n-1)}{2}\times\operatorname{cov}(X_i, X_j)$$

$$=n\times\frac{1}{n}\left(1-\frac{1}{n}\right)+n(n-1)\times\frac{1}{n^2(n-1)}$$

$$=1-\frac{1}{n}+\frac{1}{n}=1.$$

## 7.25 অনুশীলনী

7.1 দেখাও যে ধনাত্মক সম্ভাবনাযুক্ত দুটি নির্ভরতাশূন্য ঘটনা পরস্পর বিচ্ছিন্ন হতে পারে না।

7.2 দেখাও যে যদি $A$ ও $B$ দুটি পরস্পর নির্ভরতাশূন্য ঘটনা হয়, তাহলে

$$P(A \cup B) = P(A) + P(B) - P(A).\ P(B).$$

7.3 $A_1, \ldots, A_n$ যদি পরস্পর নির্ভরতাশূন্য ঘটনা হয়, তবে প্রমাণ কর যে,

$$P(\bigcup_{i=1}^{n} A_i) = 1 - \prod_{i=1}^{n} [1 - P(A_i)].$$

[ আভাস : $1 - P(\bigcup_{i=1}^{n} A_i) = P[A_1, \ldots, A_n$-এর একটিও ঘটবে না ]

$= P(A_1^*, \ldots, A_n^*$-এর সবকটিই ঘটবে )। ]

7.4 দেওয়া আছে $P(A) = \frac{3}{8}$, $P(B) = \frac{5}{8}$, $P(A \cup B) = \frac{3}{4}$ ; তা হলে $P(A|B)$ ও $P(B|A)$-এর মান কত হবে ? [ উত্তর : $\frac{2}{5}$, $\frac{2}{3}$

7.5 সমসম্ভব উপায়ে 1, 2, 3 ও 4 সংখ্যাচতুষ্টয় থেকে দুটি সংখ্যা বেে নিয়ে তাদেরকে পাশাপাশি লিখে একটি দুই অঙ্কের সংখ্যা গঠন করলে সো 2 দ্বারা বিভাজ্য হবার সম্ভাবনা কত ? [ উত্তর : $\frac{1}{2}$

7.6 সমসম্ভব উপায়ে 0, 1, 2 এবং 4-কে পরপর সাজালে প্রাপ্ত সংখ্যা 4 দিয়ে বিভাজ্য হবে এমন সম্ভাবনা কত ? [ যে সংখ্যার সর্ববামে 0 থাকে সেটিকে এই প্রসঙ্গে সংখ্যারূপে গণ্য করা চলবে না। ] [ উত্তর : $\frac{4}{9}$

7.7 দুটি অখণ্ড ধনাত্মক সংখ্যার সমষ্টি 10. তাদের গুণফল 10-এর চে বড় হওয়ার সম্ভাবনা কত ? [ উত্তর : $\frac{7}{9}$

7.8 দুটি পাত্রের প্রথমটিতে 4টি সাদা ও 2টি কালো এবং দ্বিতীয়টিতে 5 সাদা ও 2টি লাল বল আছে। চোখ বন্ধ রেখে উভয় পাত্র থেকে দুটি ক'রে ব তোলা হ'ল। মোট সংগ্রহে 2টি সাদা বল থাকবার সম্ভাবনা কত ?

[ উত্তর : $\frac{32}{10}$

7.9 দুটি পাত্রের প্রথমটিতে 3টি সাদা ও 2টি কালো এবং দ্বিতীয়টিতে সাদা ও 3টি কালো বল আছে। যদি চোখ বন্ধ রেখে প্রথমটি থেকে একটি নিয়ে দ্বিতীয় পাত্রে রেখে তারপর দ্বিতীয় পাত্র থেকে একটি বল নিয়ে প্র পাত্রে রাখা হয় এবং সবশেষে প্রথম পাত্র থেকে একটি বল তোলা হয়, ত সেটি সাদা হবে এমন সম্ভাবনা কত ? [ উত্তর : $\frac{1}{5}$

7.10 সমসম্ভব উপায়ে 1, 2, 3, 4 সংখ্যা কটি থেকে দুটি সংখ্যা বেছে নিলে তাদের সমষ্টি অযুগ্ম হবার সম্ভাবনা নির্ণয় কর যখন (i) তাদেরকে এক সঙ্গে তোলা হয়, (ii) পুনঃস্থাপনা ব্যতিরেকে একটির পর আর একটি তোলা হয় অথবা (iii) পুনঃস্থাপনাসহ একটির পর আর একটি তোলা হয়।

[ উত্তর : $\frac{2}{3}$ ; $\frac{2}{3}$ ; $\frac{1}{2}$ ]

7.11 5 জোড়া ভাইবোনের একটি গুচ্ছ থেকে সমসম্ভব উপায়ে যে কোন দু'জন বেছে নিলে তারা ভাইবোন হবার সম্ভাবনা কত? তাদের মধ্যে একটি ছেলে ও অপরটি মেয়ে হবার সম্ভাবনাই বা কত?

[ উত্তর : $\frac{1}{9}$, $\frac{5}{9}$ ]

7.12 3 জন ছেলে ও 3 জন মেয়ে একটি সারিতে বসলে

(i) মেয়ে তিনজন পাশাপাশি বসবে এমন সম্ভাবনা কত?

(ii) প্রত্যেক ছেলের পর একটি মেয়ে বা প্রত্যেক মেয়ের পর একটি ছেলে বসবার সম্ভাবনা কত? [ উত্তর : (i) $\frac{1}{5}$, (ii) $\frac{1}{10}$ ]

7.13 ধর একটি মুদ্রা নিক্ষেপ ক'রে তাতে যদি সম্মুখপার্শ্ব দেখা যায় তাহলে 1 থেকে 9 পর্যন্ত সংখ্যাগুলি থেকে একটি সংখ্যা সমসম্ভাবনাসূত্রে নেওয়া হয় এবং যদি পশ্চাৎপার্শ্ব দেখা যায় তাহলে 1 থেকে 5 পর্যন্ত সংখ্যাগুলির একটিকে সমসম্ভব উপায়ে বেছে নেওয়া হয়। তাহলে বেছে নেওয়া সংখ্যাটি যুগ্ম হবার সম্ভাবনা কত? [ উত্তর : $\frac{19}{45}$ ]

7.14 একটি শিশুপুত্র হবার সম্ভাবনা $\frac{1}{2}$ ( অর্থাৎ শিশুটি কন্যা হবার সম্ভাবনা $1-\frac{1}{2}=\frac{1}{2}$ ) ধ'রে নিয়ে বিচার কর নিম্নবর্ণিত $A_1$ ও $A_2$ ঘটনাদ্বয় পরস্পর নির্ভরতাশূন্য কিনা :—

$A_1$ : তিনজন সন্তানের একটি পরিবারে পুত্র ও কন্যা উভয় প্রকার সন্তান রয়েছে।

$A_2$ : তিনজন সন্তানের একটি পরিবারে খুব বেশী হলে একটি কন্যা রয়েছে। [ উত্তর : নির্ভরতাশূন্য হবে ]

7.15 যে কোন তিনব্যক্তির জন্মদিন সপ্তাহের তিনটি বিভিন্ন দিনে পড়বে এমন সম্ভাবনা কত ( সপ্তাহের প্রতিটি দিনকেই সমসম্ভব ধ'রে নিয়ে )?

[ উত্তর : $\frac{30}{49}$ ]

7.16 1, 2, 3, 4, 5 ও 6 সংখ্যাচিহ্নিত দুটি সুষম ছক্কা পরপর নিক্ষিপ্ত হলে তাদের ওপরে দৃষ্ট সংখ্যা-দুটির সমষ্টি 10 হবার সম্ভাবনা কত? [ উত্তর : $\frac{1}{12}$ ]

7.17 ব্রীজ খেলায় 52টি তাসকে ভালোভাবে মিশিয়ে চারজন খেলোয়াড়ের মধ্যে পৃথক্‌ভাবে বণ্টন করা হলে কোন একজন বিশেষ খেলোয়াড়ের কোন টেক্কা না পাবার সম্ভাবনা কত? [ উত্তর : $\frac{6327}{20825}$ ]

7.18 $A$ এবং $B$ দুজনে দুটি পক্ষপাতমুক্ত মুদ্রা নিয়ে বারবার উৎক্ষেপণ করতে থাকলে $B$ প্রথমবার তার মুদ্রায় 'সম্মুখপার্শ্ব' পাবার পূর্বেই $A$ তার মুদ্রায় 'সম্মুখপার্শ্ব' পাবে এমন সম্ভাবনা কত? [ উত্তর : $\frac{2}{3}$ ]

7.19 2 ইঞ্চি ব্যাসার্ধযুক্ত একটি গোলকের অভ্যন্তরে দুটি বিন্দু সমসম্ভব উপায়ে বেছে নিলে কেন্দ্রের নিকটতর বিন্দুটি কেন্দ্রের অন্ততঃ 1 ইঞ্চি দূরে পড়বার সম্ভাবনা কত? [ উত্তর : [illegible]

7.20 একটি বৃত্তের ব্যাসার্ধ যদি $r$ ইঞ্চি হয় এবং তার অভ্যন্তরে একটি বিন্দু সমসম্ভব উপায়ে নেওয়া হয়, তাহলে বৃত্তের কেন্দ্র থেকে তার দূরত্ব $x$ ইঞ্চির ($x < r$ ধ'রে) চেয়ে কম হওয়ার সম্ভাবনা কত? [ উত্তর : $\frac{x^2}{r^2}$ ]

7.21 সমস্ত প্রকৃত সংখ্যারাজি-নির্দেশক সরলরেখায় $[-2, 0]$ অন্তরের মধ্যে $b$ এবং $[0, 3]$ অন্তরের মধ্যে $a$ এই দুটি বিন্দু যদি সমসম্ভব উপায়ে গৃহীত হয়, তাহলে $b$ থেকে $a$-এর দূরত্ব 3-এর চেয়ে বেশী হবার সম্ভাবনা কত?

[ উত্তর : $\frac{1}{3}$. আভাস : সমসম্ভব উপায়ে গৃহীত $a$ ও $b$ যথাক্রমে $[a, a+da]$ ও $[b, b+db]$ অন্তরমধ্যে থাকবার সম্ভাবনা $\frac{da}{3}$ ও $\frac{db}{2}$. কাজেই $P[x < a < x+dx, y < b < y+dy] = \frac{dx}{3} \cdot \frac{dy}{2}$ ; $a-b > 3$ হতে গেলে $-2 < b < a-3$ এবং $1 < a < 3$ হতে হবে ]

7.22 একটি পাত্রে অভিন্ন আকারের 3টি সাদা এবং 2টি লাল বল আছে। যদি (1) পুনঃস্থাপনাসহ এবং (2) পুনঃস্থাপনা ব্যতিরেকে পাত্রটি থেকে একটি একটি ক'রে বল তোলা হয়, তবে প্রথমবার সাদা বল পাওয়া পর্যন্ত যতবার বল তুলতে হবে তার প্রত্যাশিত সংখ্যা কত? [ উত্তর : (1) $\frac{5}{3}$, (2) $\frac{3}{2}$ ]

7.23 একজন খেলোয়াড়কে বলা হ'ল যে, একটি ছক্কা ছুঁড়ে যদি সে কোন অযুগ্ম সংখ্যানির্দেশক চিহ্ন পায় তবে তাকে ঐ সংখ্যার দ্বিগুণ টাকা দেওয়া হবে, কিন্তু সে যদি যুগ্মসংখ্যা নির্দেশক চিহ্ন পায়, তবে তাকেই উল্টে ঐ সংখ্যার সমান টাকা দিতে হবে। এই খেলাটি কি 'ন্যায়নিষ্ঠ' হবে?

[ টীকা : একটি খেলাকে 'ন্যায়নিষ্ঠ' বলা হয় যদি ঐ খেলায় সব খেলোয়াড়ের প্রত্যাশিত লাভ বা লোকসান শূন্য হয়। ]

[ উত্তর : না, কারণ, আলোচ্য খেলোয়াড়টির প্রত্যাশিত লাভ হচ্ছে 1·00 টাকা ]

7.24 মনে কর $X$ এমন একটি সম্ভাবনা চল যে,

$$X = \begin{cases} c \text{ যদি একটি ঘটনা } A \text{ সংঘটিত হয়,} \\ d \text{ অন্যথায়।} \end{cases}$$

যদি $A$ ঘটনা সংঘটনের সম্ভাবনা হয় $p$, তবে $p, c$ ও $d$-এর মাধ্যমে $E(X)$ ও $V(X)$-কে প্রকাশ কর।

[ উত্তর : $d+(c-d)p$ ; $(c-d)^2p(1-p)$]

7.25 ধর $n$ সংখ্যক সম্ভাবনা চল $x_1, \ldots, x_n$-এর প্রত্যেকের ভেদমান $\sigma^2$ এবং তাদের প্রতি জোড়ার সহভেদমান $\rho\sigma^2$. তাহলে $V(\bar{X})$ এবং

$$E\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]\text{-এর মান কত ?}$$

[ উত্তর : $\frac{\sigma^2}{n}\{1+(n-1)\rho\}$ ; $(n-1)(1-\rho)\sigma^2$ ]

7.26 মনে কর $\{X_j\}$ হচ্ছে সমসম্ভাবনাযুক্ত পরস্পর নির্ভরতাশূন্য চলের একটি পরম্পরা যাতে $E(X_j)=0$, $V(X_j)=\sigma^2$, $j=1, 2, \ldots$. দেখাও যে, $\bar{X} \xrightarrow{P} 0$.

যদি $V(X_j)=\sigma^2\sqrt{j}$ হয়, তাহলেও কি এরকম হবে ?

[ উত্তর : হ্যাঁ। ]

7.27 মনে কর $P[X=r]=\binom{n}{r}p^r q^{n-r}$, $r=0, 1, \ldots n$ ;

$$0<p,\ q<1,\ p+q=1.$$

তাহলে $\text{cov}\left(\frac{X}{n}, \frac{n-X}{n}\right)$-এর মান নির্ণয় কর।

[ উত্তর : $\frac{-p(1-p)}{n}$ ]

7.28 একটি ছক্কা নিক্ষিপ্ত হলে তাতে যে সংখ্যানির্দেশক চিহ্ন দৃষ্ট হয় তাকে বলা যাক $X$. আবার অন্য একটি চল $Y$ নেওয়া যাক যার জন্য

$$Y = \begin{cases} +1, \text{ যদি } X \text{ যুগ্ম হয়,} \\ -1, \text{ যদি } X \text{ অযুগ্ম হয়।} \end{cases}$$

তাহলে (i) $X$, (ii) $Y$, (iii) $X+Y$, (iv) $X-Y$ ও (v) $XY$-এর সম্ভাবনা বিভাজন, গড় ও ভেদমান নির্ণয় কর।

## 7.26 নির্দেশিকা

1. Cramér, H. *The Elements of Probability Theory.* John Wiley and Sons, (1954) and Asia Publishing House.

2. Feller, W. *An Introduction to Probability Theory and its Applications,* Volume I. John Wiley and sons, 1968, Wiley Eastern 1972, and Asia Publishing House.

3. Freeman, H. *An Elementary Treatise on Actuarial Mathematics* ; Cambridge University Press, 1935.

4. Goldberg, S. *Probability : An Introduction.* Englewood Cliffs. Prentice-Hall, 1962.

5. Goon, A. M. ; Gupta, M. K. and Das Gupta, B. *Fundamentals of Statistics,* Volume I. The World Press Ltd., Calcutta, 1975.

6. Mounsey, J. *Introduction to Statistical Calculations.* English University Press, London, 1952.

7. Uspensky, J. V. *Introduction to Mathematical Probability.* McGraw Hill Book Company, Inc. New York and London, 1937.

# একচল তত্ত্বগত বিভাজন 4
## ( Univariate Theoretical Distribution )

### 4.1 ভূমিকা

পূর্ববর্তী কয়েকটি অধ্যায়ে বিস্তারিতভাবে আলোচনা করা হয়েছে কি-ভাবে কোন প্রদত্ত উপাত্তসম্ভারকে রাশিবিজ্ঞানসম্মতভাবে পর্যালোচনা ক'রে তার অন্তর্নিহিত বৈশিষ্ট্যসমূহকে বিভিন্ন মাপকের সাহায্যে স্পষ্টভাবে ফুটিয়ে তোলা যায়। সেখানে সাধারণভাবে ধ'রে নেওয়া হয়েছে যে, শুধু প্রদত্ত তথ্যাবলীই আমাদের আলোচনার বিষয়বস্তু। কিন্তু আসলে ব্যাপারটা তা নয়। রাশি-বিজ্ঞানে এটা আবশ্যিকভাবেই স্বীকৃত যে, আমরা যখন কোন প্রদত্ত রাশিতথ্য নিয়ে তার বিশ্লেষণ করতে বসি তখন আমাদের আলোচনা চূড়ান্তভাবে শুধু ঐটুকুর মধ্যেই সীমাবদ্ধ থাকে না। এই প্রসঙ্গ নিয়ে আমরা এখন একটু সবিস্তার আলোচনা করব।

রাশিবিজ্ঞানের চর্চায় যে কোন তথ্য আমাদের হাতে আসে, তাকেই অনুরূপ বহু তথ্যাবলীর একটি বৃহৎ গোষ্ঠী বা সমগ্রের একটি অংশ বা নমুনা ব'লে ধরা হয় এবং আমাদের প্রকৃত পর্যালোচনার বিষয়বস্তু ব'লে বিবেচিত হয় ঐ বৃহত্তর গোষ্ঠীটি যা সাধারণতঃ সম্পূর্ণভাবে অবেক্ষিত হয় না। আমাদের সংগ্রহে যা আছে তা ঐ বৃহত্তর গোষ্ঠীর একটি অংশমাত্র। যে কোন রাশিবিজ্ঞানবিষয়ক আলোচনার সূত্রে এই যে একটি কেবলমাত্র অংশত অবেক্ষিত তথ্যাবলীর সমগ্র গুচ্ছ বা গোষ্ঠীর কল্পনা করা হয় এবং যার বৈশিষ্ট্যাবলীর ওপর আলোকপাত করাই হচ্ছে আমাদের আলোচনা এবং পরীক্ষা-নিরীক্ষার চূড়ান্ত উদ্দেশ্য, সেই সামগ্রিক তথ্যসম্ভারকে বলা হয় পূর্ণক বা সমগ্রক (population or universe). এই সমগ্রক বা পূর্ণক সম্পর্কে ধারণা বা অনুমান-নির্ভর পর্যালোচনার উদ্দেশ্যে এর প্রতিনিধি হিসেবে আমরা এর থেকে একটি বিশেষ অংশ চয়ন ক'রে নিয়ে থাকি এবং সেই অংশের বৈশিষ্ট্যগুলি সম্পর্কে বিস্তারিত বিশ্লেষণের কাজে আমাদের প্রচেষ্টাকে নিয়োজিত করি। সমগ্রকের এরূপ প্রতিনিধিস্থানীয় অংশকে বলা হয় নমুনা বা অংশক (sample)। তারপর অংশকটির গুণধর্মগুলি আমরা সবিস্তারে বিশ্লেষণ করে থাকি, তা থেকে সমগ্রকটির অনুরূপ গুণধর্ম-সম্পর্কে অনুমান বা ধারণা করার উদ্দেশ্যে। পুরোবর্তী পরিচ্ছেদগুলিতে যে পরিসংখ্যা-বিভাজন সমুদয়ের আলোচনা করা হয়েছে বাস্তবিক সেগুলি সবই এক-একটি অংশক এবং

এদের অন্তরালে এক একটি সমগ্রক বা পূর্ণকের অস্তিত্ব বিদ্যমান রয়েছে, যদিও সে সম্পর্কে সুস্পষ্ট আলোচনা হয় নি। অংশকের আলোচনার সাহায্যে পূর্ণকের ধর্ম নিরূপণ করার উপায় হিসেবে আবিষ্কৃত পদ্ধতি হচ্ছে কতগুলি তত্ত্বগত বা ঔপপত্তিক বিভাজনের (theoretical distribution) পরিকল্পনা এবং গুণধর্মের বিস্তৃত বিশ্লেষণ। যে কোন সংগৃহীত উপাত্তকেই একটি অজ্ঞাত বা অসম্পূর্ণভাবে জ্ঞাত কোন পূর্ণকের প্রতিনিধিরূপ অংশক হিসেবে ধরা হবে এবং ঐ পূর্ণক সম্বন্ধে ধরা হবে যে, তার একটি বিভাজন আছে যাকে আমরা সম্ভাবনা তত্ত্বের সাহায্যে প্রকাশ করতে পারি। অংশক হিসাবে আলোচিত তথ্য যদি পরিমাপভিত্তিক হয়, তবে সংশ্লিষ্ট পূর্ণকের তথ্যাবলীও পরিমাপভিত্তিক হবে এবং পূর্ণকের বিভাজনটিকে একটি সম্ভাবনাশ্রয়ী চলের সাহায্যে প্রকাশ করা যাবে। এই সম্ভাবনা চলের মানগুলিই পূর্ণকের বিভিন্ন পরিমাপনযোগ্য মানগুলিকে নির্দেশ করবে। ফলে, পূর্ণকের বিভাজনটিকে একটি সম্ভাবনা চলের বিভাজনদ্বারা সূচিত করা যাবে। এরূপ সম্ভাবনাচলের বিভাজনকেই তত্ত্বগত ( অর্থাৎ সম্ভাবনা তত্ত্বগত ) বা ঔপপত্তিক বিভাজন বলা হবে। নীচে সংক্ষেপে তত্ত্বগত বিভাজনের স্বরূপটি পরিস্ফুটনের চেষ্টা করা হয়েছে।

কোন বিচ্ছিন্ন সম্ভাবনাসাপেক্ষ চল $X$ সম্পর্কে আমরা দেখেছি যে, এর মানসীমার অন্তর্গত কোন মান $x$-এর জন্যে $[X=x]$ হচ্ছে একটি ঘটনা এবং এর সম্ভাবনা $P[X=x]$-এর মান হচ্ছে সুনির্দিষ্ট। এখন $P[X=x]=f(x)$ লিখলে $f$ হচ্ছে প্রকৃত মান ধারণক্ষম একটি অপেক্ষক। এই $f$-এর দুটি প্রধান গুণধর্ম রয়েছে ; যথা :

$$\text{সব } x \text{ এর জন্যে } f(x) \geqslant 0 \qquad \cdots \quad (8.1)$$

$$\text{এবং} \quad \sum_{x} f(x)=1. \qquad \cdots \quad (8.2)$$

(8.2)-এ $x$-এর সবকটি সম্ভাব্য মানের জন্যে $f(x)$-এর মান যোগ করা হয়েছে।

এই $f$ অপেক্ষকটি $X$ চলের সম্ভাবনা বিভাজন নির্দেশ করছে [7.12. দ্রষ্টব্য ] বাস্তবিক, কোন বিচ্ছিন্ন চল $X$ এর ঔপপত্তিক বিভাজন (8.1) ও (8.2)-এ উল্লিখিত গুণধর্মবিশিষ্ট $f$ অপেক্ষকের সাহায্যে বিবৃত করা হয়।

পক্ষান্তরে, যে কোন অপেক্ষক $f$ যদি (8.1) ও (8.2) গুণধর্মবিশিষ্ট হয়, তবে বলা হয় যে, এটি একটি ঔপপত্তিক বিভাজন সূচিত করে। এস্থলে, এমন কোন সম্ভাবনাসাপেক্ষ বিচ্ছিন্ন চল $X$ আছে কি নেই যার জন্যে $P[X=x]=f(x)$,

সে সম্পর্কে আমরা উদাসীনও থাকতে পারি। যা আবশ্যিক তা হচ্ছে এই যে, $f$-এর (8.1) ও (8.2)-এ উল্লিখিত ধর্ম দুটি থাকতেই হবে এবং এই ধর্মবিশিষ্ট যে কোন অপেক্ষক $f$ একটি ঔপপত্তিক বিভাজন সূচিত করে। কিন্তু সাধারণতঃ এটা বাঞ্ছনীয় যে, (8.1) ও (8.2)-এ উল্লিখিত ধর্ম দুটি ছাড়াও $f$-এর তৃতীয় একটি গুণ থাকা উচিত যে, বাস্তবিকই যেন এমন কোন পরীক্ষণ থাকে যার ফলের ভিত্তিতে একটি বিচ্ছিন্ন চল $X$ যেন গঠন করা সম্ভব হয় যার জন্যে $P[X=x]$ এই সম্ভাবনাটিকে $f(x)$ দ্বারা সূচিত করা যায়। এরূপ একটি পরিস্থিতি যদি সত্যিই থাকে, তবে আমরা বলি যে, সেটি ঔপপত্তিক বিভাজন নির্দেশক অপেক্ষকের সম্ভাবনাভিত্তি বা সম্ভাবনা আদর্শকে (probability model) রূপায়িত করছে। এই $f(x)$-কে বিচ্ছিন্ন চল $X$-এর সম্ভাবনা ভর অপেক্ষকও (probability mass function) বলা হয়ে থাকে।

পক্ষান্তরে, ধরা যাক $X$ একটি $[\alpha, \beta]$ মানসীমা সম্পন্ন অবিচ্ছিন্ন চল যার সম্ভাবনা ঘনত্ব অপেক্ষক হচ্ছে $f$ এবং বিভাজন অপেক্ষক হচ্ছে $F(x)=\int_\alpha^x f(u)\,du$ এবং $f(x)=\frac{d}{dx}F(x)$. এই $f$ অপেক্ষকটির দুটি বিশিষ্ট ধর্ম হচ্ছে

$$\text{সব } x\text{-এর জন্যে } f(x) \geqslant 0 \tag{8.3}$$

$$\text{এবং } \int_\alpha^\beta f(x)\,dx=1. \tag{8.4}$$

এখানে বলা হবে যে, $f$ অপেক্ষকটি অবিচ্ছিন্ন চল $X$-এর তত্ত্বগত বা ঔপপত্তিক বিভাজন নির্দেশ করছে। বাস্তবিক, উপরিলিখিত (8.3) ও (8.4) ধর্মবিশিষ্ট যে কোন অপেক্ষকই কোন অবিচ্ছিন্ন চলের ঔপপত্তিক বা তত্ত্বগত বিভাজন নির্দেশ করে। অধিকন্তু এটা বাঞ্ছনীয় যে এমন একটি সম্ভাবনাসাপেক্ষ অবিচ্ছিন্ন চল $X$ থাকবে যার মানসীমা $[\alpha, \beta]$-এর কোন উপ-অন্তর $[a, b]$-এর জন্যে $P[a \leqslant X \leqslant b]$ দ্বারা নির্দেশিত সম্ভাবনার মান $\int_a^b f(x)\,dx$ এই সমাকলন সাহায্যে প্রকাশ করা যায় এবং $\int_\alpha^\beta f(x)\,dx=P[\alpha \leqslant X \leqslant \beta]=1$ হয়। যে সম্ভাবনাসাপেক্ষ ঘটনা অনুযায়ী এমন একটি সম্ভাবনাসাপেক্ষ চল $X$ নির্দেশ করা যায়, যার জন্যে $P[a \leqslant X \leqslant b]=\int_a^b f(x)\,dx$, তার সম্বন্ধে বলা হয় যে,

সেটি হচ্ছে $f$ দ্বারা নির্দেশিত তত্ত্বগত বিভাজনের সম্ভাবনা আদর্শের প্রতিভূ (probability model)।

তত্ত্বগত বিভাজনের আলোচনার সার্থকতা কী একথা স্বভাবতঃই মনে জাগে, কারণ সব পূর্ণকেরই আসল বিভাজনটি শেষ পর্যন্ত আমাদের অজ্ঞাত থেকে যেতে বাধ্য। কিন্তু যখনই কোন তত্ত্বগত বিভাজনকে আমরা বেছে নেব তা আমাদের সম্পূর্ণ জানা থাকবে এবং তা কোন অপরিজ্ঞাত পূর্ণককেই সঠিকভাবে নির্দিষ্ট করতে পারবে না। সবচেয়ে বেশী যা করতে পারে তা এই যে, নির্বাচিত তত্ত্বগত বিভাজনটি কোন প্রদত্ত পূর্ণক বিভাজনকে মোটামুটি আসন্নভাবে সূচিত করতে পারে এবং এইটুকুই আমাদের লাভ। কারণ, এই বিষয়টি আমরা কাজে লাগাতে পারি। কী ভাবে কাজে লাগানো যায় দেখা যাক। নমুনালব্ধ যে কোন রাশিতথ্যকেই আমরা একটি পূর্ণকের অনুরূপ তথ্যসম্ভারের প্রতিভূ হিসেবে দেখি। এখন এই নমুনাভিত্তিক তথ্য বিশ্লেষণ ক'রে তার কতিপয় বৈশিষ্ট্য অন্বেষণ ক'রে তার সাহায্যে একটি সুবিধামত তত্ত্বগত বিভাজন পাওয়ার চেষ্টা করা যেতে পারে, যা নমুনাটি যে পূর্ণক থেকে গৃহীত হয়েছে সেই পূর্ণকের বিভাজনটিকে যতটা সম্ভব নৈকট্য বজায় রেখে সূচিত করে। এই ব্যাপারটিকে বলা হয় প্রদত্ত নমুনাবিভাজনের সঙ্গে একটি তত্ত্বগত বিভাজনের সাযুজ্যনিরূপণ (fitting a theoretical distribution to a sample distribution)। এই বিষয়টি পরিস্ফুট করতে হলে তত্ত্বগত বিভাজনের সঙ্গে সংশ্লিষ্ট কয়েকটি বিষয়ের আলোচনা করা দরকার।

## 8.2 ঔপপত্তিক বিভাজন সংশ্লিষ্ট কয়েকটি বিষয়:

মনে কর $X$ একটি বিচ্ছিন্ন বা অবিচ্ছিন্ন সম্ভাবনা চল এবং $f$ তৎসংশ্লিষ্ট ভর বা ঘনত্ব অপেক্ষক। তাহলে,

(i) $\mu = E(X)$ হচ্ছে $X$-এর গড় বা গাণিতিক প্রত্যাশা বা যৌগিক গড়।

(ii) $\mu'_r = E(X^r)$ হচ্ছে $X$-এর $r$-তম পরিঘাত $(r=1, 2, \ldots)$।

স্পষ্টতঃই $\mu'_1 = \mu$ এবং $\mu'_0 = 1$.

(iii) $\mu'_{r:c} = E(X-c)^r$ হচ্ছে $X$-এর $r$-তম $c$-কেন্দ্রিক পরিঘাত।

লক্ষণীয় যে, $\mu'_{0:c} = 1$

(iv) $\mu_r = E(X-\mu)^r$ হচ্ছে $X$-এর $r$-তম গড়-কেন্দ্রিক পরিঘাত।

তাহলে, $\mu_0=1$, $\mu_2=\sigma^2=E(X-\mu)^2$ হচ্ছে $X$-এর ভেদমান এবং $\sigma$ হচ্ছে $X$-এর প্রমাণবিচ্যুতি। স্পষ্টতঃই $\mu_1=0$.

(v) $X^{[r]}=X(X-1)...(X-r+1)$ লিখলে

$E(X^{[r]})=E\{X(X-1)...(X-r+1)\}$-কে বলা হয় $X$-এর $r$-তম গৌণিক পরিঘাত (factorial moment)।

এখন, সহজেই দেখা যায় যে,

$$\mu_r=E(X-\mu)^r=E[X^r-\binom{r}{1}X^{r-1}\mu'_1+\binom{r}{2}X^{r-2}\mu'^2_1-\cdots]$$

$$=\mu'_r-\binom{r}{1}\mu'_{r-1}\mu'_1+\binom{r}{2}\mu'_{r-2}\mu'^2_1-\binom{r}{3}\mu'_{r-3}\mu'^3_1+\cdots$$

ফলে,
$$\left.\begin{aligned}\mu_2&=\mu'_2-\mu_1'^2,\\ \mu_3&=\mu'_3-3\mu'_2\mu'_1+2\mu'^3_1,\\ \mu_4&=\mu'_4-4\mu'_3\mu'_1+6\mu'_2\mu'^2_1-3\mu'^4_1.\end{aligned}\right\} \qquad (8.5)$$

(vi) $M_d(c)=E(|X-c|)$-কে বলা হয় $X$ এর $c$-কেন্দ্রিক চিহ্ননিরপেক্ষ গড়-বিচ্যুতি (mean deviation about $c$)।

(vii) $M_d(\mu)=E(|X-\mu|)$-কে বলা হয় $X$-এর গড়-কেন্দ্রিক চিহ্ননিরপেক্ষ গড়বিচ্যুতি। একে শুধু $M_d$ বলেও উল্লেখ করা হবে।

(viii) $X$ একটি বিচ্ছিন্ন সম্ভাবনা চল এবং $f$ তার সম্ভাবনা ভর অপেক্ষক হলে যদি

$$\sum_{x\leq M_e} f(x)=\tfrac{1}{2}=\sum_{x\geqslant M_e} f(x) \qquad \cdots \quad (8.6)$$

হয়, তবে $M_e$-কে বলা হয় $X$-এর মধ্যমমান বা মধ্যমা (median) এবং যদি

$$\frac{c}{100}=\sum_{x\leq X_c} f(x) \qquad \cdots \quad (8.7)$$

হয়, তাহলে $X_c$-কে বলে $X$-এর $c$-তম শততমক ($c^{th}$ percentile)।

তেমনি $X$ যদি সম্ভাবনা ঘনত্ব অপেক্ষক $f$ সমন্বিত এবং $[\alpha, \beta]$ মানসীমা সম্পন্ন একটি অবিচ্ছিন্ন সম্ভাবনা চল হয়, তবে

$$\int_\alpha^{M_e} f(x)\,dx=\tfrac{1}{2}=\int_{M_e}^{\beta} f(x)\,dx \qquad \cdots \quad (8.8)$$

হলে $M_e$-কে বলা হয় $X$-এর মধ্যমমান এবং

$$\frac{c}{100}=\int_a^{X_c} f(x)\,dx \qquad \cdots \quad (8.9)$$

হলে $X_c$-কে $X$-এর $c$-তম শততমক বলা হয়।

উভয়ক্ষেত্রে $c$ এর মান যথাক্রমে 25 ও 75 হলে তদনুগ $X_c$-কে $Q_1$ ও $Q_3$ সংকেতসূত্র দ্বারা নির্দেশ ক'রে তাদেরকে যথাক্রমে $X$-এর প্রথম ও তৃতীয় চতুর্থক (first and third quartile) বলা হয়। স্পষ্টতঃই $c$ এর মান 50 হলে $X_c$ হয়ে যাবে মধ্যমার সমান। এজন্যে একে দ্বিতীয় চতুর্থকও বলা যায়। যদি $d=1, 2, \ldots, 10$-এর জন্যে $c=10d$ হয়, তবে $X_c$-কে $d$-তম দশমকও ($d^{th}$ decile) বলা হয়। যদি $p=\frac{c}{100}$-এর মান 0 এবং 1-এর অন্তর্বর্তী যে কোন ভগ্নাংশ হয়, তবে তদনুযায়ী $X_c$-কে সাধারণভাবে $X$-এর $p$-তম ভগ্নাংশক ($p^{th}$ quantile or fractile) বলা হয়।

(ix) সম্ভাবনাশ্রয়ী বিচ্ছিন্ন ( অবিচ্ছিন্ন ) চল $X$-এর সম্ভাবনা ভর ( ঘনত্ব ) অপেক্ষক $f$ হলে, যে কোন $x \neq M$-এর জন্যে যদি

$$f(M) > f(x) \qquad \cdots \quad (8.10)$$

হয়, তবে $M$-কে $X$-এর ভূয়িষ্ঠক বা সম্ভাবনাগরিষ্ঠ মান (mode) বলা হয়। যদি বিচ্ছিন্ন চলের ক্ষেত্রে $x_1, x_2, \ldots, x_a$ $(x_1 \neq x_2 \neq \cdots \neq x_a)$-এর জন্যে $f(x_i)=f(M), i=1, \ldots a$, হয় এবং $X$-এর অন্য যে কোন মানের জন্যে (8.10) সত্য হয়, তবে $x_1, \ldots, x_a$-এর প্রত্যেককে $X$-এর ভূয়িষ্ঠক বলা যায়। কিন্তু এক্ষেত্রে ভূয়িষ্ঠক তাৎপর্যবিহীন হয়ে যায় এবং তখন ভূয়িষ্ঠকের অস্তিত্ব সম্বন্ধে আমাদের আগ্রহ থাকে না। যখন $X$-এর একটি মাত্র মান $M$-এর জন্যে (8.10) সত্য হবে তখনই $M$ কে আমরা $X$-এর ভূয়িষ্ঠক বলব।

অবিচ্ছিন্ন সম্ভাবনা চলের ক্ষেত্রে যদি সম্ভাবনা ঘনত্ব অপেক্ষক $f$ এমন ধর্মবিশিষ্ট হয় যে, সব $x$-এর জন্যে

$$f'(x)=\frac{d}{dx}f(x) \text{ ও } f''(x)=\frac{d^2}{dx^2}f(x)\text{-এর অস্তিত্ব থাকে,}$$

এবং যদি $\quad f'(M)=0$ ও $f''(M)<0 \qquad \cdots \quad (8.11)$

হয়, তবে $M$ হবে $X$-এর ভূয়িষ্ঠক। যদি $M$-এর একাধিক মানের জন্যে (8.11) সত্য হয়, তবে এদের প্রত্যেককে বলা হয় $X$-এর স্থানীয় ভূয়িষ্ঠক (local

mode) এবং এক্ষেত্রে $X$-কে বহুভূয়িষ্ঠক (multimodal) চল বলা হয়। যদি $M$ এর দুটি মাত্র মানের জন্যে (8.11) সত্য হয়, তবে $X$-কে দ্বিভূয়িষ্ঠকী (bimodal) চল বলা হয়।

## 8.3 কতিপয় তত্ত্বগত বিভাজন :

এখন আমরা কয়েকটি গুরুত্বপূর্ণ তত্ত্বগত বিভাজন সম্পর্কে আলোচনা করব।

### 8.3.1 বাইনোমিয়াল বিভাজন (binomial distribution) বা দ্বিপদ বিভাজন :

#### 8.3.1.1 বাইনোমিয়াল বিভাজনের সম্ভাবনা ভর অপেক্ষক ও সম্ভাবনা আদর্শ (probability mass function and probability model of binomial distribution) :

ধরা যাক $f$ এমন একটি অপেক্ষক যার জন্যে

$$f(x)=\binom{n}{x} p^x(1-p)^{n-x}=\binom{n}{x} p^x q^{n-x}, \qquad [\, q=1-p \text{ লিখে} \,]$$

$$x=0, 1, 2, \ldots, n-1, n, \qquad (8.12)$$

যেখানে $n$ একটি অখণ্ড ধনাত্মক সংখ্যা এবং $p$ হচ্ছে এমন একটি প্রকৃত সংখ্যা যে, $0 < p < 1$,

তাহলে, $f$ অপেক্ষকের সূত্রে $x$-এর বিভিন্ন মানের জন্যে $f(x)$-এর মানগুলির যে বিভাজন সূচিত হ'ল তাকে বলে বাইনোমিয়াল বিভাজন (বা দ্বিপদ বিভাজন)। স্পষ্টতঃই $x=0, 1, \ldots, n$-এর জন্যে

$$f(x) > 0 \text{ এবং } \sum_x f(x)=\sum_x \binom{n}{x} p^x q^{n-x}=(q+p)^n=1.$$

এখন, এই $f$ অপেক্ষকের সম্ভাবনা আদর্শটি (probability model) হচ্ছে নিম্নরূপ।

মনে কর $n$ সংখ্যক বেরনুলীয় প্রচেষ্টার (7.23 দ্রষ্টব্য) প্রতিটিতে 'সার্থকতা' বা 'সাফল্য' (success) লাভের সম্ভাবনা হচ্ছে $p$ এবং 'ব্যর্থতা' (failure) লাভের সম্ভাবনা হচ্ছে $q=1-p$. এখন এরকম প্রচেষ্টায় মোট $x$ সংখ্যক 'সার্থকতা' লাভের সম্ভাবনা ($x=0, 1, \ldots n-1, n$) নির্ণয়ের চেষ্টা করা যাক।

$n$ সংখ্যক বেরনুলীয় প্রচেষ্টার এমন একটি পরম্পরা কল্পনা করা যাক যার প্রথম $x$ সংখ্যক ক্ষেত্রে 'সার্থকতা' ও শেষ $(n-x)$ সংখ্যক ক্ষেত্রে 'ব্যর্থতা' ফল

দৃষ্ট হয়। এই ঘটনাকে আমরা $E$ দ্বারা চিহ্নিত করব, এবং প্রত্যেক প্রচেষ্টায় সার্থকতা ও ব্যর্থতার ঘটনাকে যথাক্রমে $S$ ও $F$ দ্বারা সূচিত করা হবে। এখন, যেহেতু প্রত্যেক প্রচেষ্টায় $S$-এর সংঘটন সম্ভাবনা $p$ এবং $F$-এর সংঘটন সম্ভাবনা $1-p=q$, এবং $S$ ও $F$ পরস্পর অনধীন ঘটনা, কাজেই

$$P(E)=p^x q^{n-x}. \qquad \cdots \quad (8.13)$$

এখন মনে করা যাক $A$ হচ্ছে সেই ঘটনা যাতে $n$-সংখ্যক বেরণুলীয় প্রচেষ্টার যে কোন পরম্পরায় মোট $x$টি সার্থকতা ও $(n-x)$টি ব্যর্থতা অবেক্ষিত হয়। তাহলে, $A$ হচ্ছে ঠিক $E$-এর মতই কতগুলি ঘটনা যাতে $x$টি $S$ এবং $(n-x)$টি $F$ ঘটনা ঘটতে দেখা যায়, কিন্তু তাদের পরম্পরাবিন্যাস বিভিন্ন। তাহলে $A$ হচ্ছে কতগুলি মৌলিক ঘটনার একটি গুচ্ছ যাতে মোট ততগুলি $E$-এর মত মৌলিক ঘটনা আছে যার সংখ্যা হচ্ছে ঠিক যতরকমভাবে মোট $x$-সংখ্যক $S$ ও $(n-x)$ সংখ্যক $F$-কে বিভিন্ন পরম্পরায় সাজানো যায়। এই সংখ্যা হচ্ছে $\frac{n!}{x!(n-x)!}=\binom{n}{x}$. আবার, স্পষ্টতঃই $A$ এর অন্তর্গত প্রত্যেকটি উপাদান হচ্ছে $E$-এর মতো এক-একটি ঘটনা যার প্রত্যেকটির সম্ভাবনা হচ্ছে $p^x q^{n-x}$. তাই, $P(A)=\binom{n}{x}p^x q^{n-x}=f(x)$. এখানে স্পষ্টতঃই $x$-এর মান $0, 1, \ldots n-1$ বা $n$ ছাড়া আর কিছু হতে পারে না। কাজেই দেখা গেল যে, বাইনোমিয়াল বিভাজন দ্বারা যে কোন নির্দিষ্ট সংখ্যক বেরণুলীয় প্রচেষ্টায় বিভিন্ন সংখ্যক সার্থকতার সম্ভাবনা নির্দিষ্ট হয়, যার ফলে দেখা গেল যে, বাইনোমিয়াল তত্ত্বগত বিভাজনের একটি বাস্তব সম্ভাবনা আদর্শ রয়েছে। উল্লেখ্য যে, বাইনোমিয়াল বিভাজন অনুসারী চল একটি বিচ্ছিন্ন চল। আরও দেখা যাচ্ছে যে, বাইনোমিয়াল বিভাজনের নির্দেশনায় $n$ ও $p$ হচ্ছে দুটি সবচেয়ে গুরুত্বপূর্ণ বিষয়। এদের মান জানা থাকলেই সম্পূর্ণ বিভাজনটি জানা যায় এবং এদেরকে পরিবর্তিত ক'রে বিভিন্ন বাইনোমিয়াল বিভাজন পাওয়া যায়। এই গুরুত্বপূর্ণ সংখ্যাদ্বয়কে বলা হয় বাইনোমিয়াল বিভাজনের বিশেষক বা নির্ণায়ক (parameter) এবং এরা যেহেতু কোন একটি সমগ্রক বা পূর্ণকের স্বরূপ প্রকাশ করে ( কারণ প্রত্যেক ঔপপত্তিক বিভাজন হচ্ছে কোন সমগ্রকের বিভাজনের প্রতিরূপ ) তাই এদেরকে পূর্ণকাঙ্কও (parameter) বলা হয়। তাহলে দাঁড়ালো যে বাইনোমিয়াল বিভাজনের দুটি পূর্ণকাঙ্ক থাকে।

### 8.3.1.2 বাইনোমিয়াল বিভাজনের পরিঘাত :

লেখা যাক, $f(x)=\binom{n}{x}p^x q^{n-x}=b(x\ ;\ n,\ p).$

সংজ্ঞানুসারে, $\mu'_r=\sum\limits_x x^r f(x).$

তাই $$\mu=\mu'_1=\sum_{x=0}^{n} x\binom{n}{x}p^x q^{n-x}$$

$$=\sum_{x=0}^{n}\frac{n\,!}{(x-1)\,!\,(n-x)\,!}\,p^x q^{n-x}$$

$$=np\sum_{x=1}^{n}\frac{(n-1)\,!}{(x-1)\,!\,(\overline{n-1}-\overline{x-1})}\ p^{x-1}.\ q^{(n-1)-(x-1)}$$

$$=np\sum_{x'=0}^{n-1}\binom{n-1}{x}p^{x'}q^{n-1-x'},\ [\ x'=x-1 \text{ লিখে } ]$$

$$=np\sum_{x'=0}^{n-1} b(x'\ ;\ n-1,\ p)=np(q+p)^{n-1}=np\ ;$$

$$\mu'_2=\sum_x x^2 f(x)=\sum_x x(x-1)f(x)+\sum_x xf(x)$$

$$=\sum_{x=0}^{n} x(x-1)\frac{n\,!}{x\,!\,(n-x)\,!}\,p^x q^{n-x}+np$$

$$=n(n-1)p^2\ \sum_{x=2}^{n}\frac{(n-2)\,!\ p^{x-2}\ q^{(n-2)-(x-2)}}{(x-2)\,!\,(\overline{n-2}-\overline{x-2})}+np$$

$$=n(n-1)p^2\ \sum_{x''=0}^{n-2}\frac{(n-2)\,!}{x''\,!\,(n-2-x'')\,!}\,p^{x''}\,q^{n-2-x''}+np$$

[ $x''=x-2$ লিখে ]

$$=n(n-1)p^2(q+p)^{n-2}+np=n(n-1)p^2+np.$$

ফলে $\mu_2=\mu'_2-\mu'^2_1=npq.$

অর্থাৎ বাইনোমিয়াল বিভাজনের গড় এবং ভেদমান যথাক্রমে $np$ এবং $npq$.

এখন, $x^3 = x(x-1)(x-2) + 3x(x-1) + x$ লিখে পাওয়া যায়

$$\mu'_{[3]} = E[X(X-1)(X-2)] = \sum_x x(x-1)(x-2)f(x)$$

$$= \sum_x x(x-1)(x-2)\frac{n!}{x!(n-x)!}p^x q^{n-x}$$

$$= n(n-1)(n-2)p^3 \sum_{x'''=0}^{n-3} \frac{(n-3)!\, p^{x'''} q^{n-3-x'''}}{x'''!\,(n-3-x''')!},$$

[ $x''' = (x-3)$ লিখে ]

$$= n(n-1)(n-2)p^3(q+p)^{n-3} = n(n-1)(n-2)p^3,$$

তেমনি, $x^4 = x(x-1)(x-2)(x-3) + 6x(x-1)(x-2) + 7x(x-1) + x$ লিখে

অনুরূপভাবে পাওয়া যায় $\mu'_{[4]} = n(n-1)(n-2)(n-3)p^4$

এবং সরল ক'রে পাই

$$\mu_3 = npq(q-p) \text{ এবং } \mu_4 = 3n^2p^2q^2 + npq(1-6pq).$$

তাই $$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(q-p)^2}{npq},\ \gamma_1 = +\sqrt{\beta_1} = \frac{q-p}{\sqrt{npq}},$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1-6pq}{npq} \text{ এবং } \gamma_2 = \beta_2 - 3 = \frac{1-6pq}{npq}.$$

তাহলে যদি $q >, =, < p$ হয়, তবে যথাক্রমে বলা যাবে যে, দ্বিপদ ( বাইনোমিয়াল ) বিভাজন দক্ষিণায়ত প্রতিবিষম, প্রতিসম ( গড় মান $= np$ এর উভয়পার্শ্বে ) ও বামায়ত প্রতিবিষম। আবার, $pq < \frac{1}{6}, =, > \frac{1}{6}$ হলে যথাক্রমে বলা হবে যে, বাইনোমিয়াল বিভাজনটি অতিতীক্ষ্ণ, সমতীক্ষ্ণ এবং স্বল্পতীক্ষ্ণ (leptokurtic, mesokurtic and platykurtic)।

### 8.3.1.3 দ্বিপদ বা বাইনোমিয়াল বিভাজনের গড়কেন্দ্রিক পরিঘাতের পৌনঃপৌনিকতাধর্ম (recursive property) :

বাইনোমিয়াল বিভাজনের $r$-তম গড়কেন্দ্রিক পরিঘাত

$\mu_r = \sum_{x=0}^{n} (x-np)^r \binom{n}{x} p^x q^{n-x}$-কে আমরা $p$ ও $n$-এর অপেক্ষক হিসেবে

গণ্য ক'রে ধ'রে নেব যে $p$ একটি অবিচ্ছিন্ন চল এবং আরও স্বীকার ক'রে নেব যে, $p$-এর সম্পর্কে $\mu_r$ এর প্রথম অন্তর্কলকের (1st derivative) অস্তিত্ব রয়েছে।

তাহলে পাওয়া যায়

$$\frac{d}{dp}\mu_r = \frac{d}{dp}\sum_{x=0}^{n}(x-np)^r \binom{n}{x} p^x q^{n-x}$$

$$= \sum_{x=0}^{n}\Big[-nr(x-np)^{r-1}p^x(1-p)^{n-x}$$

$$+x(x-np)^r p^{x-1}(1-p)^{n-x}$$

$$-(n-x)(x-np)^r p^x(1-p)^{n-x-1}\Big]\binom{n}{x}$$

$$= -nr\sum_{x=0}^{n}(x-np)^{r-1}\binom{n}{x}p^x q^{n-x}$$

$$+\sum_{x=0}^{n}\binom{n}{x}\Big[(x-np)^r p^{x-1}(1-p)^{n-x-1}(x-xp$$

$$-np+xp)\Big]$$

$$= -nr\sum_{x=0}^{n}(x-np)^{r-1}\binom{n}{x}p^x q^{n-x}+\frac{1}{pq}$$

$$\Big[\sum_{n=0}^{n}\binom{n}{x}(x-np)^{r+1}p^x q^{n-x}\Big]$$

$$= -nr\mu_{r-1}+\frac{1}{pq}\mu_{r+1}.$$

সুতরাং $$\mu_{r+1} = pq\Big[nr\mu_{r-1}+\frac{d}{dp}\mu_r\Big]. \qquad (8.14)$$

এখন, জানা আছে যে, $\mu_0 = 1$ এবং $\mu_1 = 0$. কাজেই পৌনঃপৌনিকতা সম্পর্ক (8.14) থেকে আমরা $\mu_2$, $\mu_3$, $\mu_4$ ইত্যাদি সবকটি পরিঘাত সহজেই নির্ণয় করতে পারি।

### 8.3.1.4 দ্বিপদ বা বাইনোমিয়াল বিভাজনের ভূয়িষ্ঠক :

বাইনোমিয়াল বিভাজনের সম্ভাবনা ভর অপেক্ষককে $f(x)$ লিখলে দেখা যায় যে,

$$\frac{f(x)}{f(x-1)}=\frac{n-x+1}{x}\ \frac{p}{q}.$$

ফলে, $(n-x+1)p >, =, < xq$ হলে যথাক্রমে

$f(x) >, =, < f(x-1)$ হয়,

অর্থাৎ $x <, =, > (n+1)p$ হলে যথাক্রমে

$f(x) >, =, < f(x-1)$ হয়।

কিন্তু $x$-এর মান অখণ্ড ও অ ঋণাত্মক সংখ্যা হতে হবে। কাজেই $[(n+1)p]$ অর্থাৎ $np+p$-এর চেয়ে ক্ষুদ্রতর বৃহত্তম অখণ্ড সংখ্যাই হবে বাইনোমিয়াল বিভাজনের ভূয়িষ্ঠক। এখানে, $(n+1)p$ একটি অখণ্ড সংখ্যা হলে $f(\overline{n+1}p)=f(\overline{n+1}p-1)$ হবে এবং ফলে $(n+1)p$ ও $(n+1)p-1$ উভয়কেই ভূয়িষ্ঠক বলা যাবে। কিন্তু এক্ষেত্রে সাধারণতঃ ধরা হবে যে বিভাজনটির ভূয়িষ্ঠক নেই। লক্ষণীয় যে, $np$ যদি অখণ্ড সংখ্যা হয়, তবে $np$-ই হবে ভূয়িষ্ঠক।

### 8.3.1.5 কোন প্রদত্ত নমুনালব্ধ বিভাজনের সঙ্গে একটি দ্বিপদ বা বাইনোমিয়াল বিভাজনের সাযুজ্য নিরূপণ (fitting a binomial distribution to a given sample distribution) :

ধরা যাক আমাদের হাতে এমন ধরনের কিছু উপাত্ত আছে যার প্রকৃতি থেকে এটা মনে করা যেতে পারে যে, আমরা এমন একটি পরীক্ষণ নিয়ে কাজ করছি যাতে কোন একটি ঘটনাকে দুটি মাত্র বিকল্পরূপে ঘটছে বলে দেখা যেতে পারে এবং তাতে ঐ বিকল্প রূপদুটির প্রত্যেকটি কতবার ঘটেছে বা না ঘটেছে সেসম্পর্কে বিস্তারিত তথ্য জানা আছে। এরকমক্ষেত্রে স্বভাবতঃই মনে করা যেতে পারে যে, এই তথ্যাবলী হচ্ছে এমন একটি অজ্ঞাত পূর্ণকের অংশক বা নমুনা যাকে একটি বাইনোমিয়াল বিভাজন দিয়ে মোটামুটি আসন্নভাবে রূপায়িত করা যেতে পারে। এখানে অংশকটিকে একটু বিশেষ ধরনের অংশক বলে স্বীকার ক'রে নিতে হয়। এই ধরনটি হচ্ছে এমন যে, পূর্ণকের বিভাজনটিকে একটি

বাইনোমিয়াল বিভাজন অনুসারী সম্ভাবনা চল $X$-এর বিভাজন দিয়ে নির্দিষ্ট ক'রে যদি $P[X=x]=f(x)$ লেখা হয় এবং নমুনালব্ধ মানগুলিকে $x_1, \ldots, x_i, \ldots, x_N$ দিয়ে সূচিত করা হয়, তবে ধরা হবে যে প্রত্যেক $i=1, \ldots N$-এর জন্যে $P[X=x_i]=f(x_i)$. এছাড়া নমুনালব্ধ মানগুলিকে $E=(x_1, \ldots x_i, \ldots x_N)$ দ্বারা চিহ্নিত করলে $E$-এর সম্ভাবনা ধরা হবে $P(E)=f(x_1)\ldots f(x_i)\ldots f(x_N)$ অর্থাৎ $x_1, \ldots, x_i, \ldots, x_N$-কে সমভাবে নিবেশিত $N$-সংখ্যক পরস্পর অনধীন বাইনোমিয়াল সম্ভাবনা চল $X_1, \ldots, X_i, \ldots X_N$-এর মান হিসেবে ধরা হবে। এ ধরনের নমুনাকে সমসম্ভব সরল নমুনা (simple random sample) বলা হয়। এখানে মোট নমুনা সংখ্যা হচ্ছে $N$ এবং কোন একটি মান $x$ যদি নমুনায় $f_x$ সংখ্যকবার অবেক্ষিত হয়, তবে $\frac{f_x}{N}$-কে $x$-এর নমুনালব্ধ আপেক্ষিক পরিসংখ্যা বলা হয় এবং একে $f(x)$-এর প্রাক্কলক (estimator) হিসেবে ধরা যায় এবং এই দুয়ের ঘনিষ্ঠ সম্পর্কের ভিত্তিতেই বাইনোমিয়াল বিভাজনের সাযুজ্য নিরূপণের সূত্রটি গড়ে উঠেছে। এখন লক্ষণীয় যে, পূর্ণকটি যেহেতু সাধারণতঃ সম্যকরূপে জানা থাকে না কাজেই $f(x)$-এর মধ্যে নিহিত এক বা একাধিক পূর্ণকাঙ্ক আমাদের অজ্ঞাত থাকবে। তাই সাযুজ্যনিরূপণের প্রথম ধাপে অজ্ঞাত পূর্ণকাঙ্কগুলিকে নমুনালব্ধ তথ্যের মাধ্যমে যথোপযোগী প্রাক্কলকের ব্যবহার দ্বারা অনুমান ক'রে নিতে হয়। এই প্রাক্কলনের একটি পদ্ধতি হচ্ছে পরিঘাত পদ্ধতি (method of moments)। এই পদ্ধতিটি নিম্নরূপ :

পূর্ণকনির্দেশক তত্ত্বগত বিভাজনটির সম্ভাবনা ভর অপেক্ষকে যদি $k$ সংখ্যক অজ্ঞাত পূর্ণকাঙ্ক থাকে, তবে তাদেরকে প্রথমে পূর্ণকের প্রথম $k$-সংখ্যক পরিঘাতের মাধ্যমে প্রকাশ করতে হয় এবং তারপর নমুনালব্ধ উপাত্তের ভিত্তিতে নির্ণীত প্রথম $k$-সংখ্যক পরিঘাতকে ঐ পূর্ণকের প্রথম $k$-সংখ্যক পরিঘাতের সঙ্গে যথাক্রমে সমীকরণ করা হয়। তারপর সেই $k$-সংখ্যক সমীকরণকে সমাধান ক'রে প্রদত্ত নমুনা উপাত্তের পরিঘাতের মাধ্যমে অজ্ঞাত পূর্ণকাঙ্কগুলির প্রাক্কলক নির্ধারিত হয়। এখন $f(x)$-সংশ্লিষ্ট পূর্ণকাঙ্কগুলির পরিবর্তে তাদের প্রাক্কলকগুলি ব্যবহার ক'রে $f(x)$-এর পরিবর্তে তার একটি প্রাক্কলক $\hat{f}(x)$ পাওয়া যায়। অর্থাৎ $X$ চলের সম্ভাবনা ভর অপেক্ষক $f(x)$-এ যদি $k$টি অজ্ঞাত পূর্ণকাঙ্ক $\theta_1, \ldots, \theta_k$ থাকে, তবে লেখা যায় $f(x)=f(x\,;\,\theta_1, \ldots, \theta_k)$ এবং পরিঘাত পদ্ধতি প্রয়োগে নমুনার ভিত্তিতে নমুনা উপাত্তের কোন

অপেক্ষক $\hat{\theta}_1, \ldots, \hat{\theta}_k$-কে যথাক্রমে $\theta_1, \ldots, \theta_k$-এর প্রাক্কলক হিসেবে পাওয়া গেলে $\hat{f}(x)=f(x\,;\,\hat{\theta}_1, \ldots, \hat{\theta}_k)$-কে $f(x)$-এর প্রাক্কলক বলে ধরা হয়। এখন, যেহেতু $\frac{f_x}{N}$-কে $f(x)$-এর একটি অনুমিত মান বলে ধরা যায়, তাই $f(x)$ সম্পূর্ণ জানা না থাকায় $\frac{f_x}{N}$-কে $\hat{f}(x)$-এর সঙ্গে তুলনীয় বলে মনে করা হয়। অর্থাৎ নমুনায় অবেক্ষিত $x$ মানের পরিসংখ্যা $f_x$ হচ্ছে $N\hat{f}(x)$-এর সঙ্গে তুলনীয়। এই $N\hat{f}(x)$-কে বলা হয় $x$ মানের প্রত্যাশিত পরিসংখ্যা (expected frequency). পরিঘাত পদ্ধতি সম্পর্কে দ্বাদশ পরিচ্ছেদেও কিছুটা আলোচনা করা হয়েছে।

এখন, নমুনালব্ধ বিভিন্ন $x$ মানের বাস্তব পরিসংখ্যা $f_x$ ও প্রত্যাশিত পরিসংখ্যা $N\hat{f}(x)$ মানগুলিকে পরস্পর তুলনা করলে তাদের মধ্যে যদি ভাল মিল (agreement) দেখা যায় তাহলে বলা হবে যে, প্রদত্ত নমুনাতথ্যের সঙ্গে একটি তত্ত্বগত বিভাজনের ভাল সাযুজ্য রয়েছে। অন্যথায় বলা হবে যে, আসল নমুনালব্ধ পরিসংখ্যা বিভাজনের সঙ্গে নিরূপিত তত্ত্বগত বিভাজনটির ভাল সাযুজ্য নেই অর্থাৎ অনুমিত পূর্ণকটি থেকে সম্ভবত পরিলক্ষিত নমুনাটি সংগৃহীত হয়নি।

এই পদ্ধতি অনুসরণ ক'রে কোন প্রদত্ত উপাত্তের সঙ্গে একটি বাইনোমিয়াল বিভাজনের সাযুজ্য নিরূপণ করতে গেলে আমরা দেখব যে $f(x)=\binom{n}{x}p^x q^{n-x}$-এর মধ্যে দুটি পূর্ণকাঙ্ক $n$ ও $p$ আছে। কিন্তু প্রদত্ত উপাত্তের প্রকৃতি থেকেই সাধারণতঃ $n$-এর মান জানা যায়। কাজেই একমাত্র অজ্ঞাত পূর্ণকাঙ্ক থাকে $p$. এখন $np$ হচ্ছে $f(x)$-এর বা পূর্ণকের প্রথম পরিঘাত।

আবার, $\bar{x}=\frac{1}{N}\sum_{x=0}^{n} xf_x$ হচ্ছে মানগুলির নমুনালব্ধ গড় বা প্রথম পরিঘাত।

এই দুটিকে সমান ধরলে আমরা পাই

$np=\bar{x}$ অর্থাৎ $p=\frac{\bar{x}}{n}$. কাজেই $\frac{\bar{x}}{n}$-কে $p$-এর প্রাক্কলক বলে গ্রহণ করা যায় এবং এই প্রাক্কলককে আমরা $\hat{p}$ দ্বারা চিহ্নিত করব অর্থাৎ আমরা লিখব $\hat{p}=\frac{\bar{x}}{n}$. তাহলে,

$\hat{f}(x)=f(x\ ;\ \hat{p})=\binom{n}{x}\hat{p}^{x}(1-\hat{p})^{n-x}$-কে বলা হবে $f(x)$-এর প্রাক্কলক

এবং $N\hat{f}(x)=Nf(x\ ;\ \hat{p})=N\binom{n}{x}\hat{p}^{x}(1-\hat{p})^{n-x}$ হচ্ছে

$x$-এর প্রত্যাশিত পরিসংখ্যা। বিভিন্ন $x$-এর জন্য $N\hat{f}(x)$-এর মান নির্ণয় ক'রে তাদেরকে প্রদত্ত নমুনালব্ধ $f_x$-এর মানগুলির সঙ্গে তুলনা করলেই প্রদত্ত নমুনালব্ধ উপাত্তের সঙ্গে একটি বাইনোমিয়াল বিভাজনের সাযুজ্য লক্ষ্য করা যাবে। অনেক সময় অজ্ঞাত পূর্ণকাঙ্কগুলির কোন প্রাক্কলক ব্যবহার না ক'রে তাদের যে কোন মান ধ'রে নিয়েও সাযুজ্য নিরূপণ করা যায়। এই ধ'রে নেওয়া মানটি সাধারণতঃ কোন বিচারযোগ্য প্রকল্প থেকে নেওয়া হয়। কিন্তু তাতে প্রদত্ত নমুনাটিকে প্রাক্কলনের কাজে আদৌ ব্যবহার করা হয় না। কেবলমাত্র নমুনালব্ধ পরিসংখ্যা $f_x$-এর সঙ্গে $Nf(x)$-এর তুলনা করা হয়। এদের মিল থাকলে বলা হয় যে সাযুজ্য ভাল হয়েছে। অন্যথায় বলা হয় যে সাযুজ্য ভাল হ'ল না। এখন, যে কোন সাযুজ্য কতখানি ভাল বা কতখানি মন্দ হ'ল তা বিচার ক'রে দেখবারও পদ্ধতি আছে। কিন্তু সে আলোচনা আপাততঃ স্থগিত রাখা হবে।

বাইনোমিয়াল বিভাজনের সাযুজ্য নিরূপণে $\hat{f}(x)$-এর বিভিন্ন মান নির্ণয়ের সুবিধার্থে নিম্নলিখিত বিষয়টি অনুসরণযোগ্য।

$f(x)=\binom{n}{x}p^{x}(1-p)^{n-x}$-এর জন্যে $\dfrac{f(x)}{f(x-1)}=\dfrac{n-x+1}{x}.\dfrac{p}{q},\ (x>1)$

এবং $f(0)=(1-p)^{n}$. কাজেই $f(1)=n\dfrac{p}{q}f(0), f(2)=\dfrac{n-1}{2}.\dfrac{p}{q}f(1),$

$f(3)=\dfrac{n-2}{3}.\dfrac{p}{q}f(2)$ ইত্যাদি। এখন $p$-এর পরিবর্তে $\hat{p}=\dfrac{\bar{x}}{n}$ বসিয়ে

$\hat{f}(x)$-এর মানগুলি সহজেই পাওয়া যায়।

### 8.3.2 পোয়াসঁ বিভাজন (Poisson distribution) :

#### 8.3.2.1 পোয়াসঁ বিভাজনের সম্ভাবনা ভর অপেক্ষক :

যদি একটি অপেক্ষক $f$ এমন হয় যে,

$x=0, 1, 2, \ldots\ldots$ ইত্যাদির জন্যে

$$f(x) = e^{-m}\frac{m^x}{x!}, \; m > 0,$$

তবে এরূপ $f$ দ্বারা নির্দিষ্ট বিভাজনকে পোয়াসঁ-এর বিভাজন বলা হয়। স্পষ্টতঃই,

সব $x$-এর জন্যে $f(x) > 0$

এবং $$\sum_{x=0}^{\infty} f(x) = e^{-m}\sum_{x=0}^{\infty}\frac{m^x}{x!} = e^{-m}.e^{m} = 1.$$

কাজেই $f$ দ্বারা একটি সম্ভাবনা বিভাজনকে সূচিত করা যায়। যদি কোন বিচ্ছিন্ন সম্ভাবনা চল $X$-এর সম্ভাবনা ভর অপেক্ষক $f$ উপরিউক্ত $f$-এর ন্যায় রূপবিশিষ্ট হয়, তবে $X$-কে একটি পোয়াসঁ বিভাজন অনুসারী সম্ভাবনা চল বলা হয়।

অনেক বাস্তব পরিস্থিতিতেই এমন ঘটনা ঘটে যাতে অবেক্ষিত পরিসংখ্যা-বিভাজনের গঠন ঘনিষ্ঠভাবে পোয়াসঁ বিভাজনের অনুরূপ। কাজেই পোয়াসঁ বিভাজনের সম্ভাবনা আদর্শ খুবই সহজলভ্য। কিন্তু সে আলোচনায় যাবার আগে এই বিভাজনের গাণিতিক কয়েকটি ধর্ম পরীক্ষা করা যাক।

প্রথমেই উল্লেখ্য যে পোয়াসঁ বিভাজনকে বাইনোমিয়াল বিভাজনের একটি সীমারূপ হিসেবে দেখা চলে। পূর্ণকাঙ্ক $n$ ও $p$ সম্বলিত বাইনোমিয়াল বিভাজনকে

$$b(x\,;\,n,p) = \binom{n}{x}p^x q^{n-x}, \; (q = 1-p), \text{ লিখলে}$$

$$\frac{b(x\,;\,n,p)}{b(x-1\,;\,n,p)} = \frac{n-x+1}{x}.\frac{p}{q}.$$

সুতরাং, $$\frac{b(x\,;\,n,p)}{b(x-1\,;\,n,p)} \times \frac{b(x-1\,;\,n,p)}{b(x-2\,;\,n,p)} \times \cdots \times \frac{b(1\,;\,n,p)}{b(0\,;\,n,p)}$$

$$= \frac{n-x+1}{x}.\frac{p}{q} \times \frac{n-x+2}{x-1} \times \frac{p}{q} \times \cdots \times \frac{n}{1} \times \frac{p}{q}$$

$$= \frac{(n-x+1)(n-x+2)\cdots(n-1)n}{x!}\,p^x(1-p)^x$$

অর্থাৎ $$\frac{b(x\,;\,n,p)}{b(0\,;\,n,p)} = \frac{1}{x!}(n-x+1)(n-x+2)\cdots$$

$$\cdots(n-1)n\left(\frac{m}{n}\right)^x\left(1-\frac{m}{n}\right)^x, \quad [\, m = np \text{ লিখে} \,]$$

অর্থাৎ $$b(x\,;\,n, p)=\frac{1}{x\,!}\left[\left(1-\frac{x-1}{n}\right)\left(1-\frac{x-2}{n}\right)\cdots\left(1-\frac{1}{n}\right)\cdot 1\right]m^x \times \frac{1}{\left(1-\frac{m}{n}\right)^x}\times\left(\frac{m}{n}\right)^0\left(1-\frac{m}{n}\right)^n$$

এখন, (i) $p\to 0$, (ii) $n\to\infty$, কিন্তু (iii) $np=m$ সসীম—এই তিনটি সর্তাধীনে উভয়পক্ষের সীমামান নির্ণয় ক'রে এবং সীমামানকে 'lim' সংকেত ব্যবহার ক'রে প্রকাশ করলে পাই

$$\lim b(x\,;n, p)=\frac{1}{x\,!}\left[\lim\left\{\left(1-\frac{x-1}{n}\right)\left(1-\frac{x-2}{n}\right)\cdots\left(1-\frac{1}{n}\right)1\right\}\right]m^x \times\frac{1}{\lim\left(1-\frac{m}{n}\right)^x}\times\left[\lim\left(1+\frac{-m}{n}\right)^{-\frac{n}{m}}\right]^{-m}$$

$$=e^{-m}\frac{m^x}{x\,!} \qquad \cdots \quad (8.14)$$

কারণ, $\lim\left\{\left(1-\frac{x-1}{n}\right)\left(1-\frac{x-2}{n}\right)\cdots\left(\;-\frac{1}{n}\right)\cdot 1\right\}=1,$

প্রত্যেক $x$-এর জন্য $\lim\left(1-\frac{m}{n}\right)^x=1$

এবং $\lim\left(1+\frac{-m}{n}\right)^{-\frac{n}{m}}=e\,;$

ফলে, $\left[\lim\left(1+\frac{-m}{n}\right)^{-\frac{n}{m}}\right]^{-m}=e^{-m}.$

তাহলে আমরা পেলাম পোয়াসঁ-এর তত্ত্বগত বিভাজন নির্দেশক অপেক্ষক $f(x)=e^{-m}\frac{m^x}{x\,!}$ হচ্ছে $\lim\ b(x\,;\,n, p)$ ; অর্থাৎ উল্লিখিত (i)–(iii) সর্তসমূহ-সাপেক্ষে বাইনোমিয়াল বিভাজন নির্দেশক অপেক্ষক $b(x\,;\,n, p)$-এর একটি সীমামান, বাইনোমিয়াল ও পোয়াসঁ বিভাজনসূচক অপেক্ষকদ্বয়ের মধ্যে এই যে পারস্পরিক সম্পর্কসূত্রটি পাওয়া গেল তার ব্যবহারিক মূল্য হচ্ছে এই যে, যে সকল ক্ষেত্রে কোন বিভাজনকে একটি বাইনোমিয়াল বিভাজন দ্বারা সূচিত করা যায় সেখানে যদি ঐ বাইনোমিয়াল বিভাজনের পূর্ণকাঙ্ক $p$ খুব ছোট (0-এর খুব কাছাকাছি) এবং $n$ খুব বেশী বড় হয় অথচ $np$ বেশী বড় না হয়, তবে বাইনোমিয়াল সূত্রের বিকল্পে যদি পোয়াসঁ-এর বিভাজন সূত্র অনুযায়ী কোন ঘটনার

সম্ভাবনা নির্ণয় করা হয় তাহলে উদ্ভূত ভ্রান্তির পরিমাণ হবে সামান্য। অর্থাৎ এ ক্ষেত্রে কোন নমুনাতথ্যের সঙ্গে যদি বাইনোমিয়াল বিভাজনের ভাল সাযুজ্য থাকে, তবে তার সঙ্গে পোয়াসঁ বিভাজনেরও ভাল সাযুজ্য থাকবে বলে আশা করা যেতে পারে। অথচ এতে কাজের কিছু অতিরিক্ত সুবিধা হয়। কারণ, পোয়াসঁ বিভাজনের ভর অপেক্ষকের রূপটি বাইনোমিয়াল বিভাজনের ভর অপেক্ষকের তুলনায় সরলতর।

### 8.3.2.2 পোয়াসঁ বিভাজনের পরিঘাত :

সংজ্ঞানুসারে, $\mu'_1 = \mu = E(X) = \sum_{x=0}^{\infty} x e^{-m} \frac{m^x}{x!}$

$$= e^{-m}. \sum_{x=1}^{\infty} \frac{m^x}{(x-1)!} = e^{-m}. m \sum_{x=1}^{\infty} \frac{m^{x-1}}{(x-1)!}$$

$$= e^{-m}. m \sum_{x'=0}^{\infty} \frac{m^{x'}}{x'!}, \quad [\, x' = x - 1 \text{ লিখে} \,]$$

$$= e^{-m}. m. e^m = m.$$

$$\mu'_{[2]} = E[X(X-1)] = \sum_{x=0}^{\infty} x(x-1) e^{-m} \frac{m^x}{x!}$$

$$= e^{-m} m^2 \sum_{x''=0}^{\infty} \frac{m^{x''}}{x''!}, \quad [\, x'' = x - 2 \text{ লিখে} \,]$$

$$= e^{-m}. m^2. e^m = m^2.$$

কাজেই $X^2 = X(X-1) + X$ লিখে পাওয়া যায়

$$\mu'_2 = m^2 + m, \ \mu_2 = \mu'_2 - u'^2_1 = m \text{ এবং } \sigma = +\sqrt{m}.$$

সাধারণভাবে, $\mu'_{[r]} = E[X(X-1)\cdots(X-r+1)]$

$$= \sum_{x=0}^{\infty} x(x-1)\cdots(x-r+1)\, e^{-m} \frac{m^x}{x!}$$

$$= e^{-m} m^r \sum_{x=r}^{\infty} \frac{m^{x-r}}{(x-r)!} = e^{-m} m^r \sum_{t=0}^{\infty} \frac{m^t}{t!},$$

[ $t = x - r$ লিখে ]

$$= e^{-m}. m^r. e^m = m^r.$$

তাই $X^3 = X(X-1)(X-2)+3X(X-1)+X$ লিখে পাওয়া যায়

$\mu'_3 = m^3 + 3m^2 + m$ এবং $\mu_3 = m$ ;

তেমনি $X^4 = X(X-1)(X-2)(X-3)+6X(X-1)(X-2)+7X(X-1)+X$

লিখে পাওয়া যায় $\mu'_4 = m^4 + 6m^3 + 7m^2 + m$

এবং $\mu_4 = 3m^2 + m.$

ফলে, $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{1}{m}$, $\beta_2 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1}{m}$,

$\gamma_1 = \sqrt{\beta_1} = +\sqrt{\frac{1}{m}}$, কারণ $\mu_3 = m > 0$ এবং $\gamma_2 = \beta_2 - 3 = \frac{1}{m}$.

কাজেই পোয়াসঁ-এর বিভাজন দক্ষিণায়ত প্রতিবিষম এবং অতিতীক্ষ্ণ।

### 8.3.2.3 পোয়াসঁ বিভাজনের গড়কেন্দ্রিক পরিঘাতের পৌনঃপৌনিকতাসূত্র (recurrence formula) :

সংজ্ঞানুসারে, $\mu_r = E[(X-m)^r] = \sum_{x=0}^{\infty} (x-m)^r e^{-m} \frac{m^x}{x!}.$

এখন, $\mu_r$-কে $m$-এর অপেক্ষক হিসেবে গণ্য ক'রে, $m$-কে একটি অবিচ্ছিন্ন চল ধ'রে ও $m$-এর সম্পর্কে $\mu_r$-এর অবিচ্ছিন্ন অন্তর্কলকের অস্তিত্ব স্বীকার ক'রে নিয়ে পাই

$$\frac{d}{dm}\mu_r = \sum_{x=0}^{\infty} \frac{1}{x!} \cdot \frac{d}{dm}[(x-m)^r e^{-m} m^x]$$

$$= \sum_{x=0}^{\infty} \frac{1}{x!}[e^{-m}m^x(-1)\, r\,(x-m)^{r-1} + (-1)(x-m)^r e^{-m} m^x + x e^{-m}(x-m)^r m^{x-1}]$$

$$= \sum_{x=0}^{\infty} \frac{1}{x!}\left[-re^{-m}m^x(x-m)^{r-1} + e^{-m}(x-m)^r m^x\left(\frac{x}{m}-1\right)\right]$$

$$= -r\sum_{x=0}^{\infty}(x-m)^{r-1}\frac{e^{-m}m^x}{x\,!}$$

$$+\frac{1}{m}\sum_{x=0}^{\infty}(x-m)^{r+1}e^{-m}\frac{m^x}{x\,!}$$

$$= -r\mu_{r-1}+\frac{1}{m}\mu_{r+1}$$

তাই $\mu_{r+1}=m\left[r\mu_{r-1}+\frac{d}{dm}\mu_r\right].$ ... (8.15)

এখন, জানা আছে যে, $\mu_0=1, \mu_1=0.$

তাহলে, (8.15) ব্যবহার ক'রে পোয়াসঁ বিভাজনের সবকটি গড়কেন্দ্রিক পরিঘাত সহজেই নির্ণয় করা যায়।

### 8.3.2.4 কোন প্রদত্ত বিভাজনের সঙ্গে একটি পোয়াসঁ বিভাজনের সাযুজ্য নিরূপণ (fitting a Poisson distribution to an observed distribution) :

যদি কোন উপাত্ত এমন ধরনের হয় যে তার ভিত্তিতে একটি চলের পরিসংখ্যা-বিভাজন নির্ণয় করলে এটা মনে করা যেতে পারে যে, ঐ চলটির প্রকৃত বিভাজনটি ( অর্থাৎ সমগ্র অজ্ঞাত পূর্ণকে ঐ চলটির বিভাজন ) হচ্ছে একটি পোয়াসঁ বিভাজন, তাহলে এই ধরনের উপাত্তের সঙ্গে একটি পোয়াসঁ বিভাজনের সাযুজ্য নির্ণয়ের চেষ্টা করা যেতে পারে। বাস্তবিক, এখানেও চলটির বিভাজন বাইনোমিয়াল বিভাজনের মতোই হবে। কিন্তু $p$-এর মান হতে হবে খুব কম এবং শূন্যের কাছাকাছি এবং নমুনালব্ধ গড় $\bar{x}$-এর মান যেন খুব বেশী না হয়। কারণ $\bar{x}$ হচ্ছে $np$-এর প্রাক্-কলক ( পরিঘাত পদ্ধতি অনুযায়ী ) এবং পোয়াসঁ বিভাজনের ক্ষেত্রে $np$ হচ্ছে $m$, যার মান অনত্যধিক। সাযুজ্য নিরূপণ করতে গেলে দেখা যাচ্ছে যে, পোয়াসঁ বিভাজনে একটি মাত্র অজ্ঞাত পূর্ণকাঙ্ক $m$ রয়েছে। আবার $m$ হচ্ছে E($X$) অর্থাৎ $m$ হচ্ছে চলটির সমগ্র পূর্ণকের ভিত্তিতে গঠিত গড়। যেহেতু সমগ্র পূর্ণকটি জানা নেই তাই $m$-কে নমুনাগড় $\bar{x}$-এর সমান ধরা হয় অর্থাৎ পরিঘাত পদ্ধতি অনুযায়ী $\hat{m}=\bar{x}$-কে $m$-এর

প্রাক্-কলক ধ'রে $f(x)=f(x\ ;\ m)=e^{-m}\dfrac{m^x}{x\,!}$ এর প্রাক্-কলক হিসেবে নেওয়া হবে $\hat{f}(x)=f(x\ ;\ \hat{m})=e^{-\bar{x}}\dfrac{(\bar{x})^x}{x\,!}$-কে। এখন নমুনালব্ধ মোট পরিসংখ্যা ও কোন $x$ মানের অবেক্ষিত পরিসংখ্যা যথাক্রমে $N$ ও $f_x$ হলে $\hat{f}_x$ ও $N\hat{f}(x)$ পরস্পর তুলনীয় রাশি ব'লে গণ্য হবে। এখানে একটি কথা বলা দরকার। পোয়াসঁ বিভাজনে চলটি অসীমসংখ্যক মান গ্রহণ করতে পারে। কিন্তু নমুনাটিতে তো আর আমরা অজস্র মান পর্যবেক্ষণ করি না। তার ফলে, যদিও $\sum_{x=0}^{\infty} f(x)=1$, তবু $\hat{f}(x)$-এর অবেক্ষিত মানগুলির নমুনাভিত্তিক সমষ্টি সাধারণতঃ 1-এর চেয়ে কম হবে। ধরা যাক যে, $f_x$-এর মান কেবলমাত্র $x=0, 1, \ldots, 10$-এর জন্যে দেওয়া আছে এবং অন্য $x$-এর জন্যে $f_x$ দেওয়া নেই। তাহলে $\hat{f}(x)$-এর মান $x=0, 1, \ldots, 9$ পর্যন্ত নির্ণয় ক'রে $\hat{f}(10)$ এর মান $1-\sum_{x=0}^{9}\hat{f}(x)$ ধ'রে নেওয়া যেতে পারে। তাহলে, $\sum_{x=0}^{10}\hat{f}(x)=1$ হবে। এই পদ্ধতি সার্থকভাবে কার্যকর হবে যদি প্রসঙ্গ থেকে মনে হয় যে, 10 এর চেয়ে বড় $x$-গুলির জন্যে উপাত্ত আলাদা ক'রে সংগ্রহ না ক'রে তাদের জন্যেও $x$-এর মান 10-ই ধরা হয়েছে। পরে উদাহরণ দিয়ে ব্যাপারটি আরও পরিষ্কার করার চেষ্টা করা হবে। যাই হোক তুলনার সুবিধের জন্য $\hat{f}(x)$-এর মানগুলিতে এরকম একটু সামঞ্জস্য ক'রে নিতে হয়।

এখন যদি দেখা যায় যে, $N\hat{f}(x)$ এবং $f_x$-এর মানগুলির মধ্যে বেশ ঘনিষ্ঠ মিল রয়েছে তবে আমরা বলব যে, প্রদত্ত নমুনা-বিভাজনটির সঙ্গে একটি পোয়াসঁ বিভাজনের সাযুজ্য রয়েছে। অন্যথায় আমাদের সিদ্ধান্ত বিপরীত হবে। অনেক সময় আবার $\bar{x}$-কে $m$-এর প্রাক্-কলক হিসেবে ব্যবহার না ক'রে কোন স্বেচ্ছাগৃহীত বা কোন প্রকল্প থেকে নির্ধারিত মান $m_0$-কে অজ্ঞাত $m$-এর মান হিসেবে ধ'রে নিয়েও $f(x)$-এর প্রাক্-কলক পাওয়া যেতে পারে। তাহলে আমরা $Ne^{-m_0}\dfrac{{m_0}^x}{x\,!}$-এর সঙ্গে $f_x$-এর তুলনা করব এবং যদি এদের মধ্যে সামঞ্জস্য থাকে, তবে আমরা বলব যে, সাযুজ্য ভালো পাওয়া গেছে, ইত্যাদি।

পোয়াসঁ বিভাজনের সঙ্গে কোন প্রদত্ত বিভাজনের সাযুজ্য নিরূপণে নিম্নলিখিত সম্পর্কগুলি অনুধাবনযোগ্য।

$$f(x)=e^{-m}\frac{m^x}{x\,!}\text{-এর জন্যে, } x>0 \text{ হলে } \frac{f(x)}{f(x-1)}=\frac{m}{x},$$

বা $f(x)=\frac{m}{x}f(x-1)$. এখন, $f(0)=e^{-m}$. কাজেই $f(1)=mf(0)$,

$f(2)=\frac{m}{2}f(1)$ ইত্যাদি।

### 8.3.3 অতিজ্যামিতিক বিভাজন (hypergeometric distribution) :

#### 8.3.3.1 অতিজ্যামিতিক বিভাজনের সম্ভাবনা-ভর-অপেক্ষক :

ধরা যাক, $f$ নিম্নবর্ণিত রূপবিশিষ্ট একটি অপেক্ষক :

$$f(x)=\frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}},\ 0<p<1,\quad 0<q<1,\quad p+q=1$$

$n<N$ ; $x=\max\,[0, n-Nq], 1, \ldots, \min\,[n, Np]$

[ এখানে $\min\,[n, Np]=n$, যদি $n \leqslant Np$
$\quad = Np$, যদি $n > Np.$ }

তদ্রূপ, $\max\,[0, n-Nq]=0$, যদি $n \leqslant Nq$
$\quad = n-Nq$, যদি $n, > Nq$ হয়।

আমরা লিখব $M=\min\,(n, Np)$ ]

বিস্তৃতভাবে লেখা যায়,

$$f(x)=\frac{Np\,!\;Nq\,!\;n\,!\;(N-n)\,!}{x\,!\;(Np-x)\,!\;(n-x)\,!\;(Nq-n+x)\,!\;N\,!}$$

$$=\frac{Np^{[x]}\,Nq^{[n-x]}}{N^{[n]}}\Big/ \lfloor n$$

[ উল্লেখ্য যে, $r^{[k]}=r(r-1)\cdots(r-k+1)=\frac{r\,!}{(r-k)\,!}=\binom{r}{k}\cdot k\,!$ ]

তাহলে লেখা যায়

$$f(x)=\frac{Nq^{[n]}}{N^{[n]}}\cdot\left[\frac{Np^{[x]}.n^{[x]}}{(Nq-n+x)^{[x]}}\cdot\frac{1}{x\,!}\right].$$

নিম্নলিখিত প্রসারণটিকে অতিজ্যামিতিক প্রসারণ বলা হয় :

$$F(a,b\,;\,c,t)=1+\frac{a.b}{c}\cdot\frac{t}{1\,!}+\frac{a(a+1)b(b+1)}{c(c+1)}\cdot\frac{t^2}{2\,!}+\cdots$$
$$+\frac{a(a+1)\cdots(a+x-1)b(b-1)\cdots(b+x-1)}{c(c+1)\cdots(c+x-1)}\cdot\frac{t^x}{x\,!}+\cdots$$

তাহলে, $F(a,b\,;\,c,t)$-তে $t^x$-এর সহগকে লেখা যায়

$$\frac{(-1)^x(-a)(-a-1)\cdots(-a-\overline{x-1})(-1)^x(-b)(-b-1)\cdots(-b-\overline{x-1})}{(c+x-1)(c+x-2)\cdots(c+1)c}$$

$=\dfrac{(-a)^{[x]}(-b)^{[x]}}{(c+x-1)}\cdot\dfrac{1}{x\,!}$. কাজেই $\dfrac{Np^{[x]}.n^{[x]}}{(Nq-n+x)^{[x]}}\cdot\dfrac{1}{x\,!}$ হচ্ছে অতিজ্যামিতিক প্রসারণ $F(-n,\,-Np\,;\,Nq-n+1,\,t)$-তে $t^x$-এর সহগ।

এখন লক্ষণীয় যে, $(1+t)^{Np}=\sum_x\binom{Np}{x}t^x$

ও $(1+t)^{Nq}=\sum_x\binom{Nq}{n-x}t^{n-x}$.

সুতরাং $(1+t)^{Np}.(1+t)^{Nq}=(1+t)^N$

$$=\sum_x\binom{Np}{x}t^x\cdot\sum_x\binom{Nq}{n-x}t^{n-x}$$
$$=\sum_x\left[\sum_x\binom{Np}{x}\binom{Nq}{n-x}\right]t^n.$$

সুতরাং $\sum_x\binom{Np}{x}\binom{Nq}{n-x}$ হচ্ছে $(1+t)^N$-এর প্রসারণে $t^n$-এর সহগ।

কিন্তু আবার লেখা যায় $(1+t)^N=\sum_n\binom{N}{n}t^n$.

কাজেই $\sum\binom{Np}{x}\binom{Nq}{n-x}=\binom{N}{n}$.

সুতরাং $\sum_x f(x)=\dfrac{1}{\binom{N}{n}}\sum_x\binom{Np}{x}\binom{Nq}{n-x}=\dfrac{\binom{N}{n}}{\binom{N}{n}}=1.$

আবার, স্পষ্টতঃই সব $x$-এর জন্যে $f(x) \geqslant 0$ কাজেই $f$ দ্বারা একটি সম্ভাবনা বিভাজন নির্দেশ করা যায়। বাস্তবিক, $f$ অপেক্ষক দ্বারা যে তত্ত্বগত বিভাজন সূচিত করা যায় তাকে অতিজ্যামিতিক বিভাজন বলা হয়। এখন এই ঔপপত্তিক বিভাজনের সম্ভাবনা আদর্শ কী তা দেখা যাক।

ধরা যাক, $A$ ও $B$ এই দুই প্রকারে বিভক্ত মোট $N$-সংখ্যক উপাদানবিশিষ্ট একটি পূর্ণক আছে, যার $Np$ সংখ্যক উপাদান হচ্ছে $A$ এবং বাকী $Nq$ সংখ্যক উপাদান হচ্ছে $B$. এখন মনে করা যাক যে, এই $N$ সংখ্যক উপাদান থেকে $n$টি উপাদানের একটি নমুনা বেছে নিতে হবে। এরকম মোট $\binom{N}{n}$ সংখ্যক নমুনা আছে। নমুনাটি এমনভাবে চয়ন করতে হবে যেন এই প্রত্যেকটি নমুনা নির্বাচিত হবার সম্ভাবনা সমান থাকে। তাহলে একটি নমুনা নির্বাচনকে যদি একটি পরীক্ষণ বলা হয়, তাহলে এই পরীক্ষণে মোট সমসম্ভব মৌলিক ঘটনা হচ্ছে $\binom{N}{n}$. এখন এই পরীক্ষণে অর্থাৎ নমুনা সংগ্রহে আমরা একটি ঘটনার সংঘটনে উৎসাহী। সেটি হচ্ছে এই যে, এরকম একটি নমুনায় $A$ জাতীয় উপাদান সংখ্যা হবে $x(\geqslant 0)$ বাকী $(n-x)$টি উপাদান হবে $B$. তাহলে, এই ঘটনার অনুকূলে মোট পরিস্থিতি সংখ্যা হচ্ছে $\binom{Np}{x}\binom{Nq}{n-x}$, কারণ $Np$ সংখ্যক $A$ উপাদান থেকে $x$টি $A$-উপাদান $\binom{Np}{x}$ সংখ্যক উপায়ে বেছে নেওয়া যায় এবং ঐ একই সঙ্গে $Nq$টি $B$-উপাদান থেকে $(n-x)$টি চয়ন করা যায় $\binom{Nq}{n-x}$ উপায়ে। তাহলে সম্ভাবনার পুরাতনী তত্ত্বানুসারে,

$$f(x)=\frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}}$$

হচ্ছে যথাক্রমে $Np$ ও $Nq$ সংখ্যক $A$ ও $B$ এই দুই প্রকার উপাদানসম্বলিত একটি পূর্ণক থেকে গৃহীত $n$টি উপাদানের একটি নমুনায় $x$টি $A$ ও $(n-x)$টি $B$-উপাদান নির্বাচিত হবার সম্ভাবনা। স্পষ্টতঃই এখানে নমুনাচয়ন পদ্ধতিটির বৈশিষ্ট্য হচ্ছে এই যে, এখানে সম্ভাব্য $\binom{N}{n}$ সংখ্যক নমুনার প্রত্যেকটি নির্বাচনেই সমান সম্ভাবনা আরোপ করা হয়েছে। এজাতীয় নমুনা-

চয়ন পদ্ধতিকে সরল সমসম্ভব নমুনাসংগ্রহ পদ্ধতি বলে। এখানে একটি কথা বলা অপ্রাসঙ্গিক হবে না যে, নমুনা-নির্বাচন-পদ্ধতিটি নিম্নবর্ণিতরূপে সামান্য পরিবর্তন করলেও ওপরে বর্ণিত ঘটনাটির সম্ভাবনা একই থাকবে। পদ্ধতিটি হচ্ছে এরকম :

ধরা যাক, পূর্ণক-টি থেকে প্রথমে $N$টি উপাদানের প্রত্যেকটিকে সমান সম্ভাবনা $\frac{1}{N}$ আরোপ ক'রে একটি মাত্র নমুনা নেওয়া হ'ল। তারপর বাকী $(N-1)$ সংখ্যক [ইতিপূর্বে সংগৃহীত নমুনাটিকে বাদ দিয়ে] উপাদানের প্রত্যেকটিকে সমান সম্ভাবনা $\frac{1}{N-1}$ আরোপ ক'রে দ্বিতীয় নমুনাটি গ্রহণ করা হ'ল। তারপর তৃতীয়বারে বাকী $(n-2)$ সংখ্যক উপাদান থেকে অনুরূপে একটি নমুনা নেওয়া হ'ল। এইভাবে যদি $n$-বার নমুনা সংগ্রহ করা হয় তাহলেও দেখা যায় যে, মোট $n$ উপাদানের নমুনাটিতে $x$টি $A$ ও $(n-x)$টি $B$-উপাদান পাওয়ার সম্ভাবনা হচ্ছে $\frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}}$. এই উভয়প্রকার নমুনাসংগ্রহ পদ্ধতিকেই সরল সমসম্ভব নমুনাচয়ন পদ্ধতি বলা হয় এবং আরো বিশদভাবে বলা হয় যে, এখানে নমুনাটি পুনঃপ্রত্যর্পণ ব্যতিরেকে সংগৃহীত হচ্ছে। এখানে লক্ষ্য করতে হবে যে, প্রত্যেকবার নমুনাচয়নের সঙ্গে সঙ্গে পূর্ণকটির আকৃতি ও গঠনপ্রকৃতি পরিবর্তিত হয়ে চলেছে। ফলে, এখানে যে পরীক্ষণ প্রচেষ্টাগুলি চলছে সেগুলি সম্ভাবনা তত্ত্বানুযায়ী স্বনির্ভর নয়। উদাহরণস্বরূপ বলা যেতে পারে প্রথম নির্বাচনে উপাদানটি $A$-জাতীয় হবার সম্ভাবনা হচ্ছে $\frac{Np}{N}=p$ এবং $B$-জাতীয় হবার সম্ভাবনা হচ্ছে $\frac{Nq}{N}=q$. প্রথম নির্বাচনে যদি $A$ উৎকলিত হয় (একে বলব ঘটনা $F$), তাহলে দ্বিতীয় নির্বাচনেও $A$ গৃহীত হবার (একে বলব ঘটনা $E$) সর্তাধীন সম্ভাবনা হচ্ছে $P(E|F)=\frac{Np-1}{N-1}$; কিন্তু প্রথম নির্বাচনে যদি $B$ উঠে থাকে (একে বলব ঘটনা $F^c$ অর্থাৎ $F$-এর পরিপূরক ঘটনা), তবে দ্বিতীয় নির্বাচনে $A$ সংগৃহীত হবার (ঘটনা $E$) সর্তাধীন সম্ভাবনা হচ্ছে $P(E|F^c)=\frac{Np}{N-1}$. স্পষ্টতঃই $P(E|F)\neq P(E|F^c)$. কাজেই $E$ এবং $F$

ঘটনা-দুটি অর্থাৎ প্রথম ও দ্বিতীয় নির্বাচনে $A$ জাতীয় উপাদান সংগৃহীত হবার ঘটনা-দুটি পরস্পর অনধীন নয়। তাই পরীক্ষণ প্রয়াস-দুটিও স্বনির্ভর নয়।

এখন, যদি নির্বাচন পদ্ধতিটিকে এমনভাবে পরিবর্তন করা হয় যে, প্রতিবার এক একটি ক'রে উপাদান সংগৃহীত হবার পর $A$ বা $B$ কোন্ জাতীয় তা দেখে নিয়ে সঙ্গে সঙ্গে সেটি পূর্ণক-এ ফেরৎ দিয়ে তারপর পরবর্তী নমুনাটি সংগ্রহ করা হয় অর্থাৎ যদি প্রত্যর্পণ সহযোগে নমুনাটি চয়ন করা হয়, তাহলে প্রত্যেক নির্বাচনেই পূর্ণকের আকৃতি ও গঠন-প্রকৃতি অবিকৃত থাকে এবং এভাবে নমুনা সংগ্রহের পরীক্ষণপ্রচেষ্টাগুলিকে সম্ভাবনা তত্ত্বানুযায়ী পরস্পর অনধীন বলা যায়। বাস্তবিক, এক্ষেত্রে প্রত্যেক প্রচেষ্টায় $A$ এবং $B$-এর মধ্যে একজাতীয় উপাদান পাওয়া যাবে অর্থাৎ আমরা বলতে পারি যে, যদি $A$ সংগৃহীত হয় তবে প্রচেষ্টাটি সার্থক ও যদি $B$ সংগৃহীত হয়, তবে সেটি ব্যর্থতায় পর্যবসিত হয় এবং প্রতিটি প্রচেষ্টা সার্থক ও ব্যর্থ হবার সম্ভাবনা যথাক্রমে $\frac{Np}{N}=p$ ও $\frac{Nq}{N}=q$. এক্ষেত্রে বাস্তবিক, প্রচেষ্টাগুলি বেরণুলীয় প্রচেষ্টার আকার ধারণ করে। কাজেই, স্বভাবতঃই এই নমুনাচয়ন পদ্ধতিতে সংগৃহীত $n$-সংখ্যক উপাদানে $x$টি $A$ জাতীয় উপাদান থাকার সম্ভাবনা দাঁড়াবে $g(x)=\binom{n}{x}p^x q^{n-x}$

এখন, যদি ধরা যায় যে, নমুনাচয়ন পদ্ধতিটি প্রত্যর্পণ ব্যতিরেকেই সাধিত হচ্ছে কিন্তু পূর্ণকের উপাদান সংখ্যা $N$ নমুনা উপাদানসংখ্যা $n$-এর তুলনায় খুব বড়, যার ফলে $P(E|F)=\frac{Np-1}{N-1}$, $P(E|F^c)=\frac{Np}{N-1}$ প্রভৃতি সংখ্যাগুলি কার্যতঃ $p$-এর সমান বলে গণ্য করা যেতে পারে, তবে এটা ধরা যায় যে, পরীক্ষণপ্রচেষ্টাগুলি কার্যতঃ স্বনির্ভর। তাই বলা যায় যে, এক্ষেত্রে বাইনোমিয়াল বিভাজনটিকে অতিজ্যামিতিক বিভাজনের একটি সীমারূপ হিসেবে দেখা যেতে পারে। এই ব্যাপারটি গাণিতিকভাবেও বিশ্লেষণ ক'রে দেখা যেতে পারে।

এখন, $$f(x)=\frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}}$$

$$=\frac{Np\,!}{x\,!\,(Np-x)\,!}\,\frac{Nq\,!\;n\,!\,(N-n)\,!}{(n-x)\,!\,(Nq-n+x)\,!\,N!}$$

$$= \frac{n!}{x!(n-x)!}\left[\left\{\left(\frac{Np}{N}\right)\left(\frac{Np-1}{N-1}\right)\cdots\left(\frac{Np-x+1}{N-x+1}\right)\right\}\right.$$

$$\left.\cdot\left\{\left(\frac{Nq}{N-x}\right)\left(\frac{Nq-1}{N-x-1}\right)\cdots\left(\frac{Nq-n+x+1}{N-n+1}\right)\right\}\right]$$

এখন, যদি উভয়পক্ষের সীমামান নেওয়া হয় যাতে $n << N$ অর্থাৎ $n$, $N$-এর তুলনায় অত্যন্ত ছোট, অর্থাৎ $\frac{n}{N} \simeq 0$, তাহলে আসন্নভাবে $f(x)$-এর মান দাঁড়ায় $\binom{n}{x}p^x q^{n-x} = g(x)$-এর সমান। অর্থাৎ প্রত্যেক $x$-এর জন্যে অতিজ্যামিতিক তত্ত্বগত বিভাজনের সম্ভাবনা-অপেক্ষকের সীমামান হচ্ছে বাইনোমিয়াল তত্ত্বগত বিভাজনের সম্ভাবনা-অপেক্ষকের মানের সমান।

### 8.3.3.2 অতিজ্যামিতিক বিভাজনের পরিঘাত :

$$\mu'_1 = \mu = E(X) = \sum_x xf(x)$$

$$= \sum_{x=0}^{M} x\frac{Np!}{x!(Np-x)!}\frac{Nq!\,n!\,(n-x)!}{(n-x)!\,(Nq-n+x)!\,N!}$$

$$= n\cdot\frac{Np}{N}\cdot\sum_{x=1}^{M}\frac{(Np-1)!\,Nq!\,(n-1)!\,(\overline{n-1}-\overline{x-1})!}{(x-1)!\,(Np-1-\overline{x-1})!\,(n-1-\overline{x-1})!}$$

$$\times\frac{1}{(Nq-\overline{n-1}+\overline{x-1})!\,(N-1)!}$$

$$= np\sum_{y=0}^{M-1}\frac{\binom{Np-1}{y}\binom{Nq}{n-y-1}}{\binom{N-1}{n-1}}, \quad [y = x-1 \text{ লিখে }]$$

$$= np.$$

অনুরূপে, $\mu'_{[2]} = E[X(X-1)] = \sum_x x(x-1)f(x)$

$$= n\frac{Np}{N}\cdot(n-1)\frac{Np-1}{N-1} = np\,\frac{Np-1}{N-1}\,(n-1).$$

ফলে, $\mu'_2 = E(X^2) = np\cdot\frac{Np-1}{N-1}(n-1) + np.$

কাজেই $\mu_2 = \frac{npq}{N-1}(N-n).$

অতিজ্যামিতিক বিভাজনের পূর্ণকাঙ্ক হচ্ছে তিনটি—$N$, $n$ ও $p$.

**8.3.4 সমবিভাজন বা আয়ত নিবেশন (uniform or rectangular distribution) :**

এতক্ষণ আমরা কতগুলি বিচ্ছিন্ন সম্ভাবনা চল নিয়ে আলোচনা করেছি। এবার কয়েকটি অবিচ্ছিন্ন সম্ভাবনা চলের প্রসঙ্গে আসা যাক।

ধরা যাক, $f$ হচ্ছে প্রকৃত মানাশ্রয়ী এমন একটি অপেক্ষক যার জন্যে

$$f(x)=\frac{1}{\beta-\alpha}, \quad \text{যখন } \alpha<x<\beta,$$
$$=0, \quad \text{অন্যথায়।}$$

তাহলে, যে কোন $x$-এর জন্যে $f(x)\geqslant 0$

এবং $$\int_\alpha^\beta f(x)\,dx=\frac{1}{\beta-\alpha}\int_\alpha^\beta dx=\frac{\beta-\alpha}{\beta-\alpha}=1.$$

কাজেই এই $f$ অপেক্ষকের সাহায্যে একটি সম্ভাবনা-বিভাজন নির্দেশিত করা যায়। বাস্তবিক, কোন অবিচ্ছিন্ন সম্ভাবনা চল $X$-এর মান কোন বিশেষ অন্তর $[a, b]$-এর মধ্যবর্তী $[\alpha\leqslant a\leqslant b\leqslant\beta]$ হওয়ার সম্ভাবনাকে $f$ অপেক্ষকের মাধ্যমে

$$P\left[a<X<b\right]=\int_a^b f(x)\,dx=\frac{1}{\beta-\alpha}\int_a^b dx=\frac{b-a}{\beta-\alpha}$$

আকারে প্রকাশ করা যেতে পারে। এই $f$ অপেক্ষক দ্বারা সূচিত সম্ভাবনা-বিভাজনকে সমবিভাজন (uniform distribution) বা আয়ত নিবেশন (rectangular distribution) বলা হয় এবং ওপরে যে ধরনের সম্ভাবনা চল $X$-এর উল্লেখ করা হ'ল তাকে একটি আয়ত নিবেশন সম্বলিত অবিচ্ছিন্ন সম্ভাবনা চল বলে। এ রকম নামকরণের কারণ হ'ল এই যে, $[\alpha, \beta]$ অন্তরের মধ্যগত যে কোন অন্তর $[c, d]$ নিলে, $\{[c, d]<[\alpha, \beta]$ অর্থাৎ $\alpha\leqslant c\leqslant d\leqslant\beta\}$, এর মধ্যে $X$-এর মান সীমাবদ্ধ থাকার সম্ভাবনা সর্বদা সমান হবে যদি অন্তরটির দৈর্ঘ্য সমান থাকে। কারণ,

$$P\left[c\leqslant X\leqslant d\right]=\frac{d-c}{\beta-\alpha}$$ এবং স্পষ্টতঃই এই সম্ভাবনার মান $(d-c)$-এর, অর্থাৎ $[c, d]$ অন্তরের দৈর্ঘ্যের, সমানুপাতী। একে আয়তনিবেশন বলার কারণ এই যে, যদি $f$-এর লেখ (graph) অঙ্কন করা হয়; তাহলে $(\alpha, 0)$, $(\alpha, f(\alpha))$, $(\beta, f(\beta))$ ও $(\beta, 0)$ এই চারটি সীমানানির্দেশক বিন্দু পরপর সরলরেখা দিয়ে যোগ করলে একটি আয়তক্ষেত্র পাওয়া যাবে [ চিত্র 8.1 দ্রষ্টব্য ]।

সমবিভাজনবিশিষ্ট সম্ভাবনা চল $X$-কে আমরা সমসম্ভাবনাযুক্ত সম্ভাবনা চল বলতে পারি। এর বিভাজন অপেক্ষক $F$-এর জন্যে পাওয়া যাবে

চিত্র 8.1
আয়ত নিবেশন

$$F(X)=P[X < x]=\int_{\alpha}^{x} f(x)\,dx=\frac{1}{\beta-\alpha}\int_{\alpha}^{x} dx=\frac{x-\alpha}{\beta-\alpha}.$$

স্পষ্টতঃই $\quad f(x)=\dfrac{d}{dx}\,F(x)=\dfrac{1}{\beta-\alpha}.$

এই বিভাজনের জন্যে পরিঘাত হচ্ছে

$$\mu'_1=\mu=E(X)=\int_{\alpha}^{x} xf(x)\,dx=\frac{1}{\beta-\alpha}\int_{\alpha}^{x} x\,dx=\frac{1}{\beta-\alpha}\left[\frac{x^2}{2}\right]_{\alpha}^{\beta}$$
$$=\frac{\beta^2-\alpha^2}{2(\beta-\alpha)}=\frac{\beta+\alpha}{2},$$

$$\mu'_2=E(X^2)=\int_{\alpha}^{\beta} x^2 f(x)=\frac{1}{\beta-\alpha}\int_{\alpha}^{\beta} x^2 dx$$
$$=\frac{\beta^3-\alpha^3}{3(\beta-\alpha)}=\frac{\beta^2+\alpha\beta+\alpha^2}{3}.$$

কাজেই $\quad \mu_2=\sigma^2=\mu_2-{\mu'_1}^2=\dfrac{(\beta-\alpha)^2}{12}.$

### 8.3.5 নর্ম্যাল বিভাজন (normal distribution) :

রাশিবিজ্ঞানে সর্বাধিক আলোচিত ঔপপত্তিক বিভাজন হচ্ছে নর্ম্যাল বিভাজন। এটি একটি অবিচ্ছিন্ন চলের বিভাজন। নানা কারণে রাশিবিজ্ঞানের

চর্চায় এটি একটি অত্যন্ত গুরুত্বপূর্ণ স্থান অধিকার ক'রে আছে। এখন এই বিভাজনটির গাণিতিক রূপ ও গুণধর্মাবলী আলোচনা ক'রে দেখা যাক।

### 8.3.5.1 নর্ম্যাল বিভাজনের সম্ভাবনা ঘনত্ব অপেক্ষক :

নর্ম্যাল বিভাজন সম্বলিত একটি অবিচ্ছিন্ন সম্ভাবনা চল $X$-এর সম্ভাবনা ঘনত্ব অপেক্ষকের সাধারণ গঠন প্রকৃতি হচ্ছে নিম্নরূপ :

$f(x)=c.\ exp\,[-b(x-a)^2]$ ; $-\infty < x < \infty$, $-\infty < a < \infty$, $0 < b < \infty$ ; এবং ধ্রুবক $c$-র যার মান হচ্ছে এমন যাতে

$$\int_{-\infty}^{\infty} f(x)\,dx=1 \qquad \cdots \quad (8.16)$$

এই সর্তটি পালিত হয়। এখন এই সর্তটি পালিত হওয়ার আবশ্যিক প্রয়োজন হচ্ছে $b$-র মান ধনাত্মক হওয়া। কারণ, অন্যথায়

$$\int_{-\infty}^{\infty} e^{-b(x-a)^2}dx\text{-এর}$$

মান অসীম হয়ে যাবে অর্থাৎ এই সমাকলনটি কোন সসীম মানের অভিসারী হবে না। এই $b$-এর মান সর্বদা ধনাত্মক হবে ব'লে এরপর থেকে আমরা লিখব $b=h^2$. এখন আমরা $a$ ও $b$ ( অর্থাৎ $h^2$)-কে $x$-এর পরিঘাতের মাধ্যমে প্রকাশ করব এবং $c$-র মান (8.16) থেকে নির্ণয় ক'রে $f(x)$-এর একটি সাধারণ সংহত রূপ দেওয়া হবে।

এখন, $f(x)=c.\exp\,[-h^2(x-a)^2]$.

$$\text{সুতরাং}\quad 1=\int_{-\infty}^{\infty} f(x)\,dx=c\left[\int_{-\infty}^{a}\exp[-h^2(x-a)^2]dx + \int_{a}^{\infty}\exp\,[-h^2(x-a)^2]\,dx\right]$$

$$=c\,\frac{1}{h}\left[\int_{-\infty}^{0} e^{-t^2}dt+\int_{0}^{\infty} e^{-t^2}\,dt\right],\ t=h(x-a)\ \text{লিখে}$$

$$=\frac{2c}{h}\int_{0}^{\infty} e^{-t^2}\,dt,\left[\text{কারণ,} \int_{-\infty}^{0} e^{-t^2}\,dt\text{-তে }\ t=-u\text{ লিখে}\right.$$

$$\left.\text{দেখা যায় যে,} \int_{-\infty}^{0} e^{-t^2}\,dt=\int_{0}^{\infty} e^{-t^2}\right]$$

$$=\frac{c}{h}\,\Gamma\,\tfrac{1}{2}=\frac{c}{h}\sqrt{\pi}\ ;\ [t^2=v\ \text{লিখে দেখা যায়}\,]$$

তাই $c=\frac{h}{\sqrt{\pi}}$. সুতরাং $f(x)=\frac{h}{\sqrt{\pi}}e^{-h^2(x-a)^2}$

আবার, $\mu=E(X)=\int_{-\infty}^{\infty}xf(x)\,dx=\frac{h}{\sqrt{\pi}}\int_{-\infty}^{\infty}x.\exp[-h^2(x-a)^2]dx$

$$=\frac{h}{\sqrt{\pi}}\left[\int_{-\infty}^{a}x\exp[-h^2(x-a)^2]dx+\int_{a}^{\infty}x\exp[-h^2(x-a)^2]\,dx\right]$$

$$=\frac{h}{\sqrt{\pi}}\left[\int_{-\infty}^{0}(y+a)\exp[-h^2y^2]\,dy+\int_{0}^{\infty}(y+a)\exp[-h^2y^2]dy\right],$$ $y=x-a$ লিখে

$$=\frac{h}{\sqrt{\pi}}\left[\left(\int_{-\infty}^{0}y\exp[-h^2y^2]\,dy+\int_{0}^{\infty}y\exp[-h^2y^2]dy\right)+a\left(\int_{-\infty}^{0}\exp[-h^2y^2]dy+\int_{0}^{\infty}\exp[-h^2y^2]\,dy\right)\right]$$

$$=\frac{2ah}{\sqrt{\pi}}\int_{0}^{\infty}\exp[-h^2y^2]\,dy,$$

[কারণ $y=-z$ লিখলে পাওয়া যায়

$$\int_{-\infty}^{0}y\exp[-h^2y^2]\,dy=\int_{\infty}^{0}z\exp[-h^2z^2]\,dz$$
$$=-\int_{0}^{\infty}y\exp[-h^2y^2]\,dy$$

এবং $$\int_{-\infty}^{0}\exp[-h^2y^2]\,dy=-\int_{\infty}^{0}\exp[-h^2z^2]\,dz$$
$$=\int_{0}^{\infty}\exp[-h^2y^2]\,dy\ ]$$

$=\frac{2a}{\sqrt{\pi}}\int_{0}^{\infty}e^{-t^2}dt=a$, কারণ $\int_{0}^{\infty}e^{-t^2}dt=\frac{1}{2}\Gamma\left(\frac{1}{2}\right)=\frac{\sqrt{\pi}}{2}$.

তাই লেখা যায়,

$f(x)=\frac{h}{\sqrt{\pi}}\exp[-h^2(x-\mu)^2]$ এবং এতে $\mu$ হচ্ছে $X$-এর প্রথম পরিঘাত।

এখন, $\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$

$$= \frac{h}{\sqrt{\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 \exp\left[-h^2(x-\mu)^2\right] dx$$

$$= \frac{h}{\sqrt{\pi}} \left[ \int_{-\infty}^{\mu} (x-\mu)^2 \exp\left[-h^2(x-\mu)^2\right] dx + \int_{\mu}^{\infty} (x-\mu)^2 \exp\left[-h^2(x-\mu)^2\right] dx \right]$$

$$= \frac{h}{\sqrt{\pi}} \left[ \int_{-\infty}^{0} y^2 e^{-h^2y^2} dy + \int_{0}^{\infty} y^2 e^{-h^2y^2} dy \right]$$

$$= \frac{2h}{\sqrt{\pi}} \int_{0}^{\infty} y^2 e^{-h^2y^2} dy$$

$$= \frac{2}{h^2\sqrt{\pi}} \int_{0}^{\infty} t^2 e^{-t^2} dt = \frac{1}{h^2} \cdot \frac{1}{\sqrt{\pi}} \int_{0}^{\infty} e^{-u} u^{\frac{1}{2}} du = \frac{1}{h^2} \cdot \frac{1}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right)$$

$= \frac{\Gamma(\frac{1}{2})}{2h^2\sqrt{\pi}} = \frac{1}{2h^2}$ ; ফলে, $h^2 = \frac{1}{2\sigma^2}$ , বা $h = \frac{1}{\sigma\sqrt{2}}$

তাহলে, চূড়ান্তভাবে $f(x)$ এর রূপ হ'ল

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right].$$

অনুভূমিক ও উল্লম্ব অক্ষ বরাবর যথাক্রমে $x$ এবং $f(x)$-কে সমাপন ক'রে যদি একটি লেখচিত্র আঁকা যায়, তাহলে $[x, f(x)]$ বিন্দুগুলি যোগ ক'রে যে রেখাচিত্র পাওয়া যাবে তাকে নর্ম্যাল রেখা (normal curve) বলে। $f(x)$-এর গঠন-প্রকৃতি থেকে নর্ম্যাল বিভাজনের কয়েকটি বিশিষ্ট গুণধর্ম সহজেই চোখে পড়ে। আমরা এবার সেগুলি সংক্ষেপে আলোচনা করব।

### 8.3.5.2 নর্ম্যাল বিভাজনের বা নর্ম্যাল রেখার ধর্ম :

1. সাধারণভাবে নর্ম্যাল বিভাজনের সম্ভাবনা-ঘনত্ব-অপেক্ষকের রূপ হচ্ছে

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp.\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right], \quad -\infty < x < \infty,$$

$$-\infty < \mu < \infty, 0 < \sigma < \infty.$$

এখানে $\mu$ হচ্ছে বিভাজনটির গাণিতিক প্রত্যাশা ও $\sigma$ হচ্ছে তার প্রমাণবিচ্যুতি। স্পষ্টতঃই $\mu$ এবং $\sigma$-এর বিভিন্ন মানের জন্যে বিভিন্ন নর্ম্যাল বিভাজন পাওয়া যাবে এবং $\mu$ ও $\sigma$ জানা থাকলেই একটি নর্ম্যাল বিভাজনকে সম্পূর্ণভাবে

নির্দেশিত করা যাবে। এইজন্যে $\mu$ ও $\sigma$-কে নর্ম্যাল বিভাজনের দুটি পূর্ণকাঙ্ক বলা হয়।

2. $\mu$ থেকে সমদূরবর্তী যে কোন দুটি মান $\mu \pm \delta$-এর জন্যে $f(\mu-\delta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\delta^2}{2\sigma^2}}$ ও $f(\mu+\delta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\delta^2}{2\sigma^2}}$. ফলে, সব $\delta$-এর জন্যে $f(\mu-\delta) = f(\mu+\delta)$ ; অর্থাৎ $f(x)$ রেখা $\mu$-এর উভয়পার্শ্বে প্রতিসম।

3. নর্ম্যাল বিভাজনের মধ্যমমান, ভূয়িষ্ঠক ও গাণিতিক গড় অভিন্ন।

ধরা যাক, $X$ হচ্ছে $\mu$ ও $\sigma$ পূর্ণকাঙ্কদ্বয় সম্বলিত একটি নর্ম্যাল চল।

$$\text{তাহলে,}\quad P[X < \mu] = \int_{-\infty}^{\mu} f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\mu} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{t^2}{2}} dt, \quad \left[\frac{x-\mu}{\sigma} = t \text{ লিখে}\right]$$

$$= \frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{u^2}{2}} du, \quad [\, t = -u \text{ লিখে}\,]$$

$$= \tfrac{1}{2} \cdot \frac{1}{\sqrt{\pi}} \Gamma(\tfrac{1}{2}) = \tfrac{1}{2} \;;$$

$$\text{সুতরাং}\quad P[X > \mu] = 1 - P[X < \mu] = 1 - P[\bar{x} < \mu],$$

[ কারণ X অবিচ্ছিন্ন চল ],

$= \frac{1}{2}$, অর্থাৎ $\mu$ হচ্ছে $X$-এর মধ্যমমান।

আবার, $X$-এর মান $\mu$ থেকে যতই দূরে সরে যাবে $\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$-এর মান ততই বাড়বে ; ফলে, $e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$ বা $f(x)$-এর মান ততই কমবে এবং $X$-এর মান $\mu$ এর যতই কাছাকাছি হবে $f(x)$ ততই বাড়বে। অর্থাৎ $X$-এর মান $\mu$-এর সমান হলে $f(x)$ মান গরিষ্ঠ হবে। কাজেই $\mu$ হচ্ছে $f(x)$-এর ভূয়িষ্ঠক। ফলে,

$f(\mu) = \frac{1}{\sigma\sqrt{2\pi}}$ হচ্ছে $f(y)$-এর সর্বোচ্চ মান। তাই $f(x)$ রেখার সর্বোচ্চ কোটি হচ্ছে $\frac{1}{\sigma\sqrt{2\pi}}$.

4. নর্ম্যাল বিভাজনের গড়কেন্দ্রিক যে কোন বিযুগ্ম পরিঘাতের মান হচ্ছে শূন্য।

নর্ম্যালরেখার প্রতিসাম্য ধর্মের জন্যেই এরকম হবে।

$$\mu_{2r+1} = E(X-\mu)^{2r+1}$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^{2r+1} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\, dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \left[ \int_{-\infty}^{\mu} (x-\mu)^{2r+1} e^{-\frac{1}{2\sigma^2}(x-\mu^2)}\, dx \right.$$

$$\left. + \int^{\infty} (x-\mu)^{2r+1} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\, dx \right]$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \left[ \int^{0} t^{2r+1} e^{-\frac{t^2}{2\sigma^2}}\, dt + \int_0^{\infty} t^{2r+1} e^{-\frac{t^2}{2\sigma^2}}\, dt \right],$$

[$t = (x-\mu)$ লিখে ]

$$= \frac{1}{\sigma\sqrt{2\pi}} \left[ (-1)^{2r+1} \int_0^{\infty} u^{2r+1} e^{-\frac{u^2}{2\sigma^2}}\, du + \int_0^{\infty} u^{2r+1} e^{-\frac{u^2}{2\sigma^2}} du \right]$$

$= 0.$ [ প্রথম সমাকলকে $u = -t$ এবং দ্বিতীয়টিতে $u = t$ লিখে ]

কিন্তু, $\mu_{2r} = \int_{-\alpha}^{\infty} (x-\mu)^{2r} f(x)\, dx$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^{2r} \exp.\left[ -\frac{(x-\mu)^2}{2\sigma^2} \right] dx$$

$$= \frac{2}{\sigma\sqrt{2\pi}} \int_0^{\infty} t^{2r} \exp.\left[ -\frac{t^2}{2\sigma^2} \right] dt$$

[ ঠিক আগের মতো ধাপে ধাপে এগিয়ে ]

$$= \frac{1}{\sqrt{\pi}}\, 2^r \sigma^{2r} \int_0^{\infty} e^{-w} w^{r-\frac{1}{2}}\, dw, \qquad \left[ w = \frac{t^2}{2\sigma^2} \text{ লিখে} \right]$$

$$= \frac{2^r \sigma^{2r}}{\sqrt{\pi}} \Gamma(r+\tfrac{1}{2}) = (2r-1)(2r-3) \quad \cdots \quad 5.3.1 \ \sigma^{2r}$$

$$= \frac{(2r)!}{2^r . r!} \sigma^{2r}.$$

উদাহরণস্বরূপ, $\mu_2 = \sigma^2$, $\mu_4 = 3\sigma^4$, $\beta_1 = 0$, $\beta_2 = 3$, $\gamma_1 = 0$, $\gamma_2 = 0$.

5. নর্ম্যাল রেখার আকৃতি সম্পর্কে ইতিমধ্যেই কিছু ইঙ্গিত দেওয়া হয়েছে। প্রথমতঃ এটি অবশ্যই একটি ঘণ্টাকৃতিবিশিষ্ট রেখা (bell-shaped curve). $\mu$ বিন্দুতে সর্বোচ্চ মান $\frac{1}{\sigma\sqrt{2\pi}}$ গ্রহণ করার পর $\mu$-এর উভয়পার্শ্বে প্রতিসাম্য বজায় রেখে $x$-এর মান বাড়ার সঙ্গে সঙ্গে $f(x)$-এর মান ক্রমশঃ কমতে কমতে শূন্যের কাছাকাছি চ'লে যেতে থাকে। এখন, $f(x)$-এর অন্তর্কলন নিয়ে দেখা যায়

$$f'(x) = \frac{d}{dx} f(x) = \frac{1}{\sigma\sqrt{2\pi}} \left\{ -\frac{1}{\sigma^2}(x-\mu) \right\} \exp.\left[ -\tfrac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right]$$

$$= \frac{-1}{\sqrt{2\pi}} \cdot \frac{(x-\mu)}{\sigma^3} \exp.\left[ -\frac{(x-\mu)^2}{2\sigma^2} \right]$$ এবং ফলে $f'(\mu) = 0$

এবং $$f''(x) = \frac{d^2}{dx^2} f(x) = -\frac{1}{\sigma^3\sqrt{2\pi}} \exp\left[ -\tfrac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right]$$

$$-\frac{1}{\sqrt{2\pi}} \cdot \frac{(x-\mu)^2}{\sigma^5} \cdot \exp.\left[ -\tfrac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right]$$

$$= \frac{-1}{\sqrt{2\pi}} \frac{1}{\sigma^3} \left\{ 1 - \left(\frac{x-\mu}{\sigma}\right)^2 \right\} \exp.\left[ -\frac{1}{2\sigma^2}(x-\mu)^2 \right].$$

কাজেই, $f''(\mu) = \frac{-1}{\sigma^3\sqrt{2\pi}} < 0$, যেহেতু $\sigma > 0$.

সুতরাং, $x = \mu$-তে $f(x)$-এর একটি স্থানীয় চরম মান রয়েছে। এইজন্যে $\mu$ হচ্ছে $f(x)$-এর ভূয়িষ্ঠক।

এখন, $f''(x) = 0$ সমীকরণ থেকে পাই $x = \mu \pm \sigma$ অর্থাৎ $\mu - \sigma$ এবং $\mu + \sigma$ বিন্দু-দুটি হচ্ছে নর্ম্যাল রেখার দুটি নতিবিন্দু (points of inflexion). এদের তাৎপর্য হচ্ছে এই যে $\mu \pm \sigma$-এর মধ্যবর্তী $x$-এর মানসমূহের জন্যে $f(x)$ রেখার আকৃতি হচ্ছে অবতল (concave) এবং $\mu \pm \sigma$-এর বহিঃস্থিত $x$-এর মান-সমুদয়ের জন্যে $f(x)$ রেখার আকৃতি হচ্ছে উত্তল (convex) ধরনের।

6. যদি একটি অবিচ্ছিন্ন সম্ভাবনা চল $X$-এর $\mu$ গড় ও $\sigma$ প্রমাণবিচ্যুতি

সম্বলিত সম্ভাবনা-বিভাজন নর্ম্যাল প্রকৃতির হয়, তাহলে সংক্ষেপে লেখা হয় যে $X$-এর বিভাজন হচ্ছে

$$N(\mu, \sigma^2).$$

ধরা যাক, $\tau = \frac{X-\mu}{\sigma}$. তাহলে, $\tau$ একটি সম্ভাবনা চল হবে এবং $X$-এর সম্ভাবনা-বিভাজন থেকেই $\tau$-এর সম্ভাবনা-বিভাজন নির্ণয় করা সম্ভব। প্রকৃতপক্ষে, $\tau$-এর বিভাজন হচ্ছে $N(0, 1)$. কারণ, $X$-এর সম্ভাবনা ঘনত্ব অপেক্ষক হচ্ছে

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \text{ exp. } \left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$$

এবং $X$-এর সম্ভাবনা উপাদান (probability element) হচ্ছে

$$dF(x) = P[x - \tfrac{1}{2}dx \leqslant X \leqslant x + \tfrac{1}{2}dx]$$

$$= f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \cdot \text{ exp. } \left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] dx.$$

সুতরাং $\tau$-এর সম্ভাবনা উপাদান হচ্ছে

$$dG(t) = P[t - \tfrac{1}{2}dt \leqslant \tau \leqslant t + \tfrac{1}{2}dt]$$

$$= \frac{1}{\sqrt{2\pi}} \text{ exp. } \left(-\frac{t^2}{2}\right) dt = \phi(t)\, dt \text{ ( লেখা যাক )।}$$

তাহলে, $\phi(t) = \frac{1}{\sqrt{2\pi}} \cdot \text{ exp. } \left(-\frac{t^2}{2}\right)$ হচ্ছে $\tau$-এর সম্ভাবনা-ঘনত্ব-অপেক্ষক। স্পষ্টতঃই $\phi(t)$ হচ্ছে 0 গড় ও 1 প্রমাণবিচ্যুতি বিশিষ্ট নর্ম্যাল বিভাজনের সম্ভাবনা-ঘনত্ব-অপেক্ষক।

$\tau$-এর বিভাজন-অপেক্ষক হচ্ছে

$$\Phi(t) = P[\tau \leqslant t] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \text{exp. } \left(-\frac{y^2}{2}\right) dy = \int_{-\infty}^{t} \phi(y)\, dy.$$

এখন, $t = \frac{x-\mu}{\sigma}$ লিখে পাই

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \text{ exp. } \left[-\frac{1}{2\sigma^2}(x-\mu)^2\right] = \frac{1}{\sigma}\,\frac{1}{\sqrt{2\pi}} \text{ exp. } \left(-\frac{t^2}{2}\right)$$

$$= \frac{1}{\sigma}\phi(t) = \frac{1}{\sigma}\,\phi\left(\frac{x-\mu}{\sigma}\right), \qquad \cdots \quad (8.18)$$

$$\text{এবং } F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp.\left[-\frac{1}{2\sigma^2}(u-\mu)^2\right] du$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} \exp.\left(-\frac{t^2}{2}\right) dt = \Phi\left(\frac{x-\mu}{\sigma}\right). \quad \cdots \quad (8.19)$$

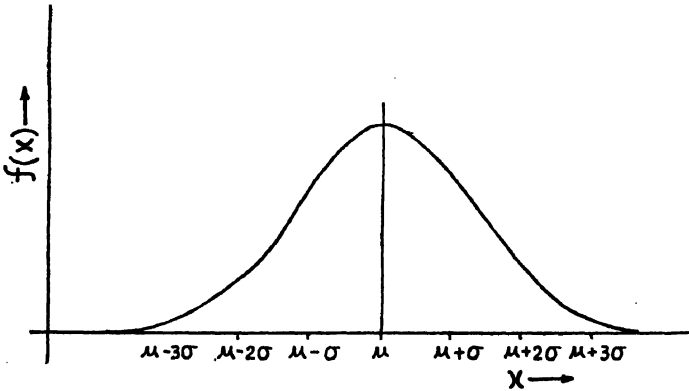
(8.18) ও (8.19) সম্পর্ক দুটি খুবই প্রয়োজনীয়। কারণ, $\phi$ ও $\Phi$ অপেক্ষক-দুটির অতি সরল গঠনপ্রকৃতি থেকে স্পষ্টই বোঝা যাচ্ছে যে, $t$-এর বিভিন্ন মানের জন্যে $\phi(t)$ ও $\Phi(t)$-এর মানগুলি অতি সহজেই নির্ণয় করা যায়। এখন যদি অন্য যে কোন গড় $\mu$ ও প্রমাণবিচ্যুতি $\sigma$ সম্বলিত কোন নর্ম্যাল বিভাজনের জন্যে $f$ ও $F$-এর যে কোন বিন্দুতে মান নির্ণয় করা প্রয়োজন হয়, তবে সে প্রয়োজন আমরা খুব সহজেই মেটাতে পারি (8.18) ও (8.19)-এ উল্লিখিত সম্পর্ক দুটি কাজে লাগিয়ে। যদি $x$ বিন্দুতে $f$ ও $F$-এর মান জানতে হয়, তবে $t = \frac{x-\mu}{\sigma}$ অর্থাৎ $x = \mu + t\sigma$ লিখে (8.18) ও (8.19) থেকে নির্ণেয় মান-দুটি অতি সহজেই বের করা যায়। এই উদ্দেশ্যে সবচেয়ে সুবিধেজনক ব্যবস্থা হচ্ছে অনেকগুলি $t$-এর জন্যে $\phi$ ও $\Phi$-এর মান বের ক'রে সারণীভুক্ত ক'রে রাখা [ বাস্তবিক, E. S. Pearson ও H. O. Hartley কর্তৃক সঙ্কলিত Biometrika Tables for Statisticians, Vol. I-এ এগুলি সারণীভুক্ত রয়েছে ] এবং কোন অন্তর্বর্তী মানের জন্যে $\phi$ ও $\Phi$-এর মান প্রয়োজন হলে অন্তঃপ্রক্ষেপণ (interpolation) পদ্ধতি প্রয়োগ করা [ এই পুস্তকের শেষাংশে সংযোজিত পরিশিষ্ট অংশে অন্তঃপ্রক্ষেপণ নীতি ও তার প্রয়োগ সম্পর্কে বিস্তারিত আলোচনা করা হয়েছে ]। কাজেই $N(0, 1)$ বিভাজনটি খুবই গুরুত্বপূর্ণ। একে অনেকসময় প্রমাণ বা সমক নর্ম্যাল বিভাজন (standard normal distribution) বলা হয়। তেমনি $N(0, 1)$ বিভাজনবিশিষ্ট সম্ভাবনা চল $\tau$-কেও নর্ম্যাল বিভাজন তত্ত্বে একটি বিশেষ মর্যাদা দেওয়া হয়েছে এবং একে বলা হয় প্রমাণীকৃত নর্ম্যাল বিভেদ চল বা মৌল নর্মাল চল (standardised normal deviate), কারণ অন্য যে কোন গড় $\mu$ ও প্রমাণবিচ্যুতি $\sigma$ বিশিষ্ট নর্ম্যাল সম্ভাবনা চল $X$ থাকলে তার থেকে গড় $\mu$ বিয়োগ ক'রে ও বিয়োগফলকে $\sigma$ দিয়ে ভাগ ক'রে যে চল পাওয়া যায় তাই হচ্ছে $\tau$.

7. উল্লিখিত মৌল নর্ম্যাল চলের সারণী থেকে সহজেই দেখা যায় যে,

$$P[\mu - 3\sigma \leqslant X \leqslant \mu + \sigma] = P[-3 \leqslant \tau \leqslant 3]$$

$$= 0{\cdot}997 \text{ ( আসন্নভাবে )।}$$

এথেকে বোঝা যায় যে, নর্ম্যাল রেখাতলবর্তী আয়তনের ( যার মোট পরিমাণ হচ্ছে 1, কারণ $\int_{-\infty}^{\infty} f(x)\,dx = 1$) প্রায় সবটুকুই ( মোটামুটি 0·997 তিন দশমিক স্থান পর্যন্ত আসন্ন ), $\mu - 3\sigma$ থেকে $\mu + 3\sigma$ পর্যন্ত বিস্তৃত $x$-মানের অন্তরমধ্যে আবদ্ধ। এজন্যে $(\mu - 3\sigma,\ \mu + 3\sigma)$-কে নর্ম্যাল চল $X$-এর বা নর্ম্যাল বিভাজনের 'কার্যকর প্রসার' (effective range) বলা হয়। এর অর্থ হচ্ছে এই যে, যদিও $X$-এর আসল প্রসার হচ্ছে অসীম ($-\infty$ থেকে $+\infty$), কিন্তু ব্যবহারিক ক্ষেত্রে এই প্রসারকে $\mu - 3\sigma$ থেকে $\mu + 3\sigma$ পর্যন্ত বিস্তৃত ব'লে ধ'রে নেওয়া যায়। কারণ এই সীমার বহিঃস্থ $x$ মানের জন্যে নর্ম্যাল রেখা-তলবর্তী আয়তনের পরিমাণ নগণ্য, 1000 ভাগের মধ্যে প্রায় 3 ভাগ মাত্র। সংক্ষেপে বলা হয় যে, $\mu$ থেকে উভয় পার্শ্বে $3\sigma$ সীমার মধ্যেই $f(x)$ রেখার মুখ্যভাগ (0·997 ভাগটি ) বিধৃত থাকে।

এই আলোচনা থেকে নর্ম্যাল রেখার আকৃতি সম্পর্কে যথেষ্ট স্পষ্ট ধারণা করা যায়। নীচের ছবিটি ( চিত্র 8.2 ) দেখলে এই ধারণা আরও স্পষ্ট হবে।

চিত্র 8.2
নর্ম্যাল নিবেশন

8. ধরা যাক, $Q_1$ ও $Q_3$ যথাক্রমে $N(0, 1)$ বিভাজনবিশিষ্ট সম্ভাবনা চল $X$-এর প্রথম ও তৃতীয় চতুর্থক। তাহলে,

$$\frac{1}{4} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Q_1} e^{-\frac{t^2}{2}}\, dt,$$

[ স্পষ্টতঃই, $Q_1 < 0$, কারণ $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{y^2}{2}} dy$ এই অপেক্ষকটি $t$-এর সঙ্গে সঙ্গে ক্রমাগত বেড়েই চলে এবং $\Phi(0) = \frac{1}{2}$ কারণ $\tau$-এর মধ্যমমান হচ্ছে 0 ]

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{t^2}{2}} dt - \frac{1}{\sqrt{2\pi}} \int_{Q_1}^{0} e^{-\frac{t^2}{2}} dt$$

$$= \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_{Q_1}^{0} e^{-\frac{t^2}{2}} dt.$$

সুতরাং, $$\frac{1}{\sqrt{2\pi}} \int_{Q_1}^{0} e^{-\frac{t^2}{2}} dt = \frac{1}{4}. \quad \cdots \quad (8.20)$$

আবার, $$\frac{3}{4} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Q_3} e^{-\frac{t^2}{2}} dt$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{t^2}{2}} dt + \frac{1}{\sqrt{2\pi}} \int_{0}^{Q_3} e^{-\frac{t^2}{2}} dt$$

[ লক্ষণীয় যে, $Q_3 > 0$ ]

$$= \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_{0}^{Q_3} e^{-\frac{t^2}{2}} dt.$$

সুতরাং, $$\frac{1}{\sqrt{2\pi}} \int_{0}^{Q_3} e^{-\frac{t^2}{2}} dt = \frac{1}{4}. \quad \cdots \quad (8.21)$$

এখন, (8.20) ও (8.21) সম্পর্ক-দুটি ব্যবহার ক'রে এবং নর্ম্যাল রেখার প্রতিসাম্য ধর্ম থেকে পাওয়া যায় $Q_1 = -Q_3$. তাছাড়া, Biometrika Tables, Vol. I থেকে পাই $Q_3 = \cdot 67$ (প্রায়) কারণ $\Phi(\cdot 67) \simeq 0\cdot 75$. কাজেই $Q_1 = -\cdot 67$. এখন, $\cdot 75 = \Phi(\cdot 67) = \Phi(Q_3) = F(\mu + \sigma Q_3) = F(\mu + \cdot 67\sigma)$, $\mu$ ও $\sigma$ হচ্ছে যথাক্রমে অপর কোন নর্ম্যাল চল $X$-এর গড় ও প্রমাণবিচ্যুতি। কাজেই $N(\mu, \sigma^2)$-এর তৃতীয় চতুর্থক হচ্ছে $\mu + \cdot 67\sigma$ এবং স্পষ্টতঃই প্রথম চতুর্থক $Q_1$ হচ্ছে $\mu - \cdot 67\sigma$.

সুতরাং চতুর্থক বিচ্যুতি হচ্ছে $\frac{Q_3 - Q_1}{2} = \cdot 67\sigma$.

9. $\phi$ ও $\Phi$ অপেক্ষকের গুণধর্ম :

আমরা জানি $\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ ; তাই $\phi(-t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} = \phi(t)$, $\cdots$ (8.22)

অর্থাৎ, $\phi$ হচ্ছে 0-এর উভয়পার্শ্বে প্রতিসম। এছাড়া,

$$\Phi(t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{u^2}{2}}\,du \text{ ; কাজেই } \Phi(-t)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-t} e^{-\frac{u^2}{2}}\,du$$

$$=\frac{1}{\sqrt{2\pi}}\int_{\infty}^{t} e^{-\frac{u^2}{2}}\,(-du), \text{ [ } u=-t \text{ লিখে ]}$$

$$=\frac{1}{\sqrt{2\pi}}\int_{t}^{\infty} e^{-\frac{u^2}{2}}\,du$$

$$=\frac{1}{\sqrt{2\pi}}\left[\int_{-\infty}^{\infty} e^{-\frac{u^2}{2}}\,du-\int_{-\infty}^{t} e^{-\frac{u^2}{2}}\,du\right]$$

$$=\Phi(\infty)-\Phi(t)=1-\Phi(t) \quad \cdots \quad (8.23)$$

(8.22) ও (8.23) সম্পর্ক-দুটি থাকার ফলে, $t$-এর কেবলমাত্র ধনাত্মক মানের জন্যে $\phi(t)$ ও $\Phi(t)$-এর মান জানা থাকলেই কাজ চলে এবং সেজন্যেই Biometrika Tables for Statisticians, Vol I-এ সেগুলিই কেবল লিপিবদ্ধ আছে। [ উদাহরণতঃ উল্লিখিত সারণী থেকে $\Phi(3)-\Phi(-3)=2\Phi(3)-1=0{\cdot}997$ ( আসন্নভাবে ) ]

10. $P[\mu-\in\leqslant X\leqslant\mu+\in]=\frac{1}{2}$ —এই সম্বন্ধটিতে উল্লিখিত $\in$-কে বলা হয় $X$-এর সম্ভাব্য ভ্রান্তি। $X$-এর বিভাজন নর্ম্যাল হলে আমরা পাই

$$\frac{1}{2}=P\left[-\frac{\in}{\sigma}\leqslant\frac{X-\mu}{\sigma}\leqslant\frac{\in}{\sigma}\right]=P\left[-\frac{\in}{\sigma}\leqslant\tau\leqslant\frac{\in}{\sigma}\right]$$

$$=P\left(\tau\leqslant\frac{\in}{\sigma}\right)-P\left(\tau\leqslant-\frac{\in}{\sigma}\right)=\Phi\left(\frac{\in}{\sigma}\right)-\Phi\left(-\frac{\in}{\sigma}\right)$$

$$=2\Phi\left(\frac{\in}{\sigma}\right)-1.$$

অর্থাৎ $\Phi\left(\frac{\in}{\sigma}\right)=\frac{3}{4}$. তাই $\frac{\in}{\sigma}=0{\cdot}67$ ( প্রায় )

অর্থাৎ $\in=0{\cdot}67\sigma$.

### 8.3.5.3 নর্ম্যাল বিভাজনের সঙ্গে কোন প্রদত্ত পরিসংখ্যা বিভাজনের সাযুজ্য নিরূপণ :

একটি নর্ম্যাল বিভাজনের রূপপ্রকৃতি ও গঠনবৈচিত্র্য আমরা দেখলাম। সচরাচর আমরা অবিচ্ছিন্ন চলের মানের ভিত্তিতে যে সমস্ত তথ্য নিয়ে তাদের বৈশিষ্ট্য পর্যালোচনা ক'রে থাকি সেগুলি বিশ্লেষণ করলে অধিকাংশ ক্ষেত্রেই দেখা যাবে যে, তাদের ধরন প্রায়ই এমন যে, তাদের ভিত্তিতে গড়া পরিসংখ্যা-

বিভাজনের স্বরূপ অনেকটা নর্ম্যাল বিভাজনের অনুরূপ। প্রদত্ত পরিসংখ্যা-বিভাজনটির জন্যে যদি আয়তচিত্র (histogram) এঁকে নেওয়া যায়, তবে তার চেহারা থেকেই অনেকটা আঁচ ক'রে নেওয়া যাবে তার আকৃতি অনেকটা নর্ম্যাল বিভাজনের রেখাচিত্রের অনুরূপ কি না। যেমন, যদি দেখা যায় যে, আয়তচিত্রের ঠিক মধ্যবর্তী আয়তক্ষেত্রটির উচ্চতা সর্বাধিক ও তার উভয়পার্শ্বস্থ আয়তগুলির উচ্চতা ধীরে ধীরে মোটামুটি প্রতিসাম্য বজায় রেখে কমতে কমতে সর্বশেষ প্রান্তীয় আয়তদ্বয়ের উচ্চতা খুব কম হয়ে পড়ে এবং প্রায় অনুভূমিক অক্ষের সমীপবর্তী হয়ে পড়ে, তবে আশা করা যায় যে, চিত্রটিতে আয়তশীর্ষের মধ্যবিন্দুগুলি যোগ ক'রে যদি একটি মসৃণ ( অবন্ধুর ) অবিচ্ছিন্ন রেখা টানা যায়, তবে তা অনেকটা একটি নর্ম্যাল রেখার আকার ধারণ করে। যদি এরকম পরিস্থিতির উদ্ভব হয়, তবে ধ'রে নেওয়া হয় যে, প্রদত্ত নমুনাভিত্তিক পরিসংখ্যা-বিভাজনটি যে পূর্ণক থেকে নেওয়া হয়েছে সেটিকে কোন নর্ম্যাল বিভাজন $N(\mu, \sigma^2)$ দ্বারা সম্পূর্ণভাবে সূচিত করা যায়। অর্থাৎ প্রদত্ত পরিসংখ্যা-বিভাজনটি একটি অবিচ্ছিন্ন সম্ভাবনা চল $X$-এর পূর্ণক বিভাজনের প্রতিনিধিত্ব করছে, যে বিভাজনটি হচ্ছে একটি নর্ম্যাল বিভাজন $N(\mu, \sigma^2)$. এরপর পরিঘাত পদ্ধতি অনুসরণ ক'রে অজ্ঞাত পূর্ণকাঙ্কদ্বয় $\mu$ ও $\sigma$-এর প্রাক্কলক হিসেবে নেওয়া হয় যথাক্রমে অবেক্ষিত বিভাজনের ভিত্তিতে নির্ণীত নমুনা গড় $\bar{x}$ ও নমুনা প্রমাণ বিচ্যুতি $s$-কে। এরপর $X$-এর সম্ভাবনা ঘনত্ব অপেক্ষক $f(x)=f(x\,;\,\mu, \sigma)$-এর আসন্ন মান হিসেবে $\hat{f}(x)=f(x\,;\,\bar{x}, s)=\dfrac{1}{s\sqrt{2\pi}}\exp\left[-\dfrac{1}{2s^2}(x-\bar{x})^2\right]$-কে ব্যবহার ক'রে প্রদত্ত বিভাজনে বিবেচিত বিভিন্ন শ্রেণী অন্তরগুলির মধ্যে $X$-এর মান সীমাবদ্ধ থাকার সম্ভাবনার আসন্নমান নির্ণয় করা হয়। এই সম্ভাবনাগুলিকে মোট পরিসংখ্যা $n$ দিয়ে গুণ ক'রে ঐ শ্রেণীঅন্তরগুলির প্রত্যাশিত পরিসংখ্যা নির্ণয় করা হয়। সবশেষে এদের সঙ্গে অবেক্ষিত পরিসংখ্যাগুলিকে তুলনা করা হয়। এইভাবে সাযুজ্য নিরূপণের পর আরও এক ধাপ এগিয়ে যাওয়া যায়। এটি হচ্ছে একটি লেখভিত্তিক পদ্ধতি অনুসরণ। এতে প্রদত্ত পরিসংখ্যা-বিভাজনের ভিত্তিতে অঙ্কিত আয়তচিত্রটির ওপর একটি যথোপযুক্ত নর্ম্যাল রেখা সংস্থাপনের চেষ্টা করা হয়। এই উদ্দেশ্যে আগের মতোই প্রদত্ত পরিসংখ্যা-বিভাজনটিকে একটি নর্ম্যাল বিভাজনের $N(\mu, \sigma^2)$-এর প্রতিনিধি ব'লে ধ'রে $f(x)=f(x\,;\,\mu, \sigma)$ ব্যবহার করা হয়। তারপর শ্রেণীপার্থক্য $h$ সমন্বিত শ্রেণীঅন্তর

$\left(x-\frac{h}{2},\ x+\frac{h}{2}\right)$-এর অবেক্ষিত পরিসংখ্যা $f_x$ থেকে পরিসংখ্যা ঘনত্ব (frequency density) $p_x=\frac{f_x}{h}$ নির্ণয় ক'রে একে $n\hat{f}(x)$ এর সঙ্গে তুলনীয় ব'লে ধরা হয়। এখন, সুবিধেমতো $x$-এর কতগুলি মান বেছে নিয়ে তাদের জন্যে $n\hat{f}(x)$ এর-মান নির্ণয় ক'রে পূর্বে অঙ্কিত চিত্রটির [ যার আয়তক্ষেত্রগুলির উচ্চতা হচ্ছে শ্রেণীগুলির পরিসংখ্যা ঘনত্বের সমান ] ওপরই যদি $x, n\hat{f}(x)$-এর লেখটি আঁকা যায় তবে $(x, n\hat{f}(x))$ বিন্দুগুলি একটি হস্তাঙ্কিত রেখার সাহায্যে অবিচ্ছিন্ন ও যথাসম্ভব মসৃণভাবে যোগ করলে একটি রেখাচিত্র পাওয়া যাবে যা হচ্ছে প্রদত্ত বিভাজনটির সঙ্গে সাযুজ্যরক্ষাকারী একটি নর্ম্যাল রেখা। এই বিশ্লেষণে (8.22) ও (8.23) সম্বন্ধ-দুটি খুব গুরুত্বপূর্ণ। এদের প্রয়োগমূল্য সহজেই দেখা যেতে পারে। ধরা যাক, শ্রেণীসীমান্তের ভিত্তিতে নির্দিষ্ট $(x_1, x_2)$ একটি শ্রেণীঅন্তর। তাহলে,

$$P=\mathrm{P}[x_1 \leqslant X \leqslant x_2]=\int_{x_1}^{x_2} f(x)\,dx\text{-এর প্রাক্কলক হিসেবে নেব}$$

$$\hat{P}=\int_{x_1}^{x_2} \hat{f}(x)\,dx=\int_{x_1}^{x_2} f(x\,;\,\bar{x},\,s)dx=\frac{1}{s\sqrt{2\pi}}\int_{x_1}^{x_2} e^{-\frac{1}{2s^2}(x-\bar{x})^2}\,dx.$$

এখন, $t=\frac{x-\bar{x}}{s}$ লিখলে $\hat{P}=\frac{2}{\sqrt{2\pi}}\int_{\frac{x_1-\bar{x}}{s}}^{\frac{x_2-\bar{x}}{s}} \exp\left(-\frac{t^2}{2}\right)dt$

$$=\Phi\left(\frac{x_2-\bar{x}}{s}\right)-\Phi\left(\frac{x_1-\bar{x}}{s}\right).$$

এখন, Biometrika Tables, Vol. I থেকে যে কোন $x_1, x_2, \bar{x}, s$-এর জন্যে $\Phi(t)$ ও তার থেকে $\hat{P}$-এর মান নির্ণয় করা অতি সহজ কাজ। অবশ্য এজন্যে অন্তঃপ্রক্ষেপণ পদ্ধতি অনুসরণ করা প্রায় সবসময়ই দরকার হয় এবং তখন ঋজুরৈখিক অন্তঃপ্রক্ষেপণ পদ্ধতি অনুসরণ করাই প্রচলিত রীতি। এখন, $(x_1, x_2)$ শ্রেণীটির প্রত্যাশিত পরিসংখ্যা হচ্ছে

$$n\,\hat{P}=n\left[\Phi\left(\frac{x_2-\bar{x}}{s}\right)-\Phi\left(\frac{x_1-\bar{x}}{s}\right)\right].$$

আবার, $\hat{f}(x)=\frac{1}{s\sqrt{2\pi}}\exp\left[-\frac{1}{2s^2}(x-\bar{x})^2\right]$

এখন $t=\frac{x-\bar{x}}{s}$ লিখলে

$$\hat{f}(x)=\frac{1}{s}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}=\frac{\phi(t)}{s}=\frac{\phi\left(\frac{x-\bar{x}}{s}\right)}{s}$$

এবং $x$ বিন্দুতে সাযুজ্য নির্ধারণকারী নর্ম্যাল রেখার কোটি হচ্ছে $n\hat{f}(x)=\frac{n}{s}\phi\left(\frac{x-\bar{x}}{s}\right)$. এটি হচ্ছে প্রদত্ত পরিসংখ্যা বিভাজনের ভিত্তিতে অঙ্কিত আয়তচিত্রস্থিত $(x_2-x_1)$ ভূমিযুক্ত আয়তক্ষেত্রটির উচ্চতা অর্থাৎ $(x_1, x_2)$ শ্রেণীর পরিসংখ্যা ঘনত্বের সঙ্গে তুলনীয়।

নর্ম্যাল বিভাজনের দুটি বিশেষ প্রয়োগক্ষেত্র খুব গুরুত্বপূর্ণ। মনে কর $X$ একটি বাইনোমিয়াল বিভাজন-সম্বলিত বিচ্ছিন্ন চল। এখন, যদি $P[x_1 \leqslant X \leqslant x_2]$-এর মান নির্ণয় করতে হয়, তাহলে অনেক সময় হয়ত অনেকগুলি $x$-এর জন্যে $b(x\,;\,n, p)=\binom{n}{x}p^x q^{n-x}$-এর মান বের ক'রে তাদের সমষ্টি নির্ণয় করতে হতে পারে। কিন্তু $n$-এর মান বেশী বড় হলে এই মান নির্ণয় কষ্টসাধ্য হয়ে পড়ে। তখন দু-একটি সাধারণ সর্তসাপেক্ষে এই অসুবিধা কিছুটা কমানো যায়।

$$\delta_1=\frac{x_1-np}{\sqrt{npq}},\ \delta_2=\frac{x_2-np}{\sqrt{npq}},\ \ h=\frac{1}{\sqrt{npq}}$$

লিখলে যদি $n$ খুব বড় হয় এবং $p$ ও $q$ খুব ছোট না হয়, [ আরও সঠিকভাবে বলতে গেলে যদি $\lim h\delta_1{}^3=0$ ও $\lim\limits_{n\to\infty} h\delta_2{}^3=0$ হয় ], তবে

প্রমাণ করা যায় যে, $\sum\limits_{x=x_1'}^{x_2} b(x\,;\,n, p)=\sum\limits_{x=x_1}^{x_2}\binom{n}{x}p^x q^{n-x}$-এর মান হচ্ছে

$\Phi\left(\frac{x_2-np}{\sqrt{npq}}+\frac{1}{2}\right)-\Phi\left(\frac{x_1-np}{\sqrt{npq}}-\frac{1}{2}\right)$-এর খুব কাছাকাছি। অর্থাৎ বাইনোমিয়াল বিভাজনকে এক্ষেত্রে নর্ম্যাল বিভাজন দ্বারা পরিবর্তিত করলে খুব ভ্রান্তি হয় না অর্থাৎ $\frac{X-np}{\sqrt{npq}}$-এর সম্ভাবনা-বিভাজন আসন্নভাবে নর্ম্যাল বিভাজনের [ অর্থাৎ $N(0, 1)$-এর ] অনুরূপ, যদিও $X$ নিজে একটি বিচ্ছিন্ন চল।

আরও দেখানো যায় যে, $\delta = \frac{x-np}{\sqrt{npq}}$ লিখলে, যদি $\lim_{n\to\infty} \frac{\delta}{n} = 0$ হয়, তবে

$$b(x\,;\,n,\,p) = \binom{n}{x} p^x q^{n-x} \text{ এবং}$$

$$\frac{1}{\sqrt{2\pi npq}} \exp\left[-\frac{(x-np)^2}{2npq}\right]\text{-এর মধ্যে পার্থক্য খুব কম,}$$

অর্থাৎ $\lim\left[b(x\,;\,n,p) - \frac{1}{\sqrt{2\pi npq}} \exp\left\{-\frac{(x-np)^2}{2npq}\right\}\right] = 0.$

অনেকটা তেমনিভাবে, $X$ যদি পোয়াসঁ বিভাজন-সম্বলিত একটি সম্ভাবনা চল হয় ও তার পূর্ণকাঙ্ক $m$ খুব বড় হয়, তাহলে দেখানো যায় যে

$$\sum_{x=x_1}^{x_2} P(x\,;\,m) = \sum_{x=x_1}^{x_2} e^{-m}\frac{m^x}{x\,!}\text{-এর মানও}$$

$\Phi\left(\frac{x_2-m}{\sqrt{m}} + \frac{1}{2}\right) - \Phi\left(\frac{x_1-m}{\sqrt{m}} - \frac{1}{2}\right)$ এর মানের খুব কাছাকাছি অর্থাৎ এক্ষেত্রে পোয়াসঁ বিভাজনের সীমারূপ হচ্ছে নর্ম্যাল বিভাজন। সঠিক অর্থে $\frac{X-m}{\sqrt{m}}$-এর সম্ভাবনা-বিভাজনের সীমারূপ হচ্ছে নর্ম্যাল বিভাজন $N(0, 1)$.

### 8.3.5.4 রাশিবিজ্ঞানে নর্ম্যাল বিভাজনের গুরুত্ব সম্পর্কে সংক্ষিপ্ত আলোচনা :

ওপরে আমরা নর্ম্যাল বিভাজন সম্পর্কে অনেকটা বিস্তারিত আলোচনা করেছি। এর কারণ হচ্ছে এই যে, রাশিবিজ্ঞানের চর্চায় এটি একটি অত্যন্ত গুরুত্বপূর্ণ স্থান অধিকার ক'রে আছে। কী কী কারণে এর এই বিশেষ মর্যাদা, তা একটু খতিয়ে দেখা যাক।

প্রথমতঃ, মূলতঃ অন্তর্সম (basically homogeneous) কোন পূর্ণক থেকে যদি কোন রাশিতথ্যের নমুনা নেওয়া হয়, তাহলে তার ভিত্তিতে গঠিত পরিসংখ্যা-বিভাজনের প্রকৃতি প্রায়ই নর্ম্যাল বিভাজনের অনুরূপ হতে দেখা যায়। এজন্যে হাতে কোন রাশিতথ্য থাকলে অনেকসময়ই ধ'রে নেওয়া হয় যে, এটি কোন নর্ম্যাল পূর্ণক থেকে গৃহীত নমুনাবিশেষ। এই অঙ্গীকারের কতগুলি সুবিধেজনক ফলশ্রুতি আছে। যেমন, নর্ম্যাল পূর্ণক থেকে গৃহীত নমুনা থেকে গঠিত অনেক নমুনাঙ্কের সম্ভাবনা-বিভাজন খুব সহজেই নির্ণয় করা যায় এবং

তাদের ভিত্তিতে মূল পূর্ণকটির পূর্ণকাঙ্ক সম্পর্কে প্রাক্কলন, প্রকল্পবিচার ইত্যাদি নানাপ্রকার অনুমান-ক্রিয়া সম্পাদন খুব সহজ হয়ে পড়ে। অবশ্য মূল অঙ্গীকারটি সত্য না হলে এ সমস্ত অনুমান ভ্রান্ত হয়ে পড়বে।

এ ছাড়া আর একটি আবিষ্কার নর্ম্যাল বিভাজনকে সবচেয়ে বেশী গুরুত্ব দিয়েছে। সেটি এই যে, মূল পূর্ণকটির নিবেশন যাই হোক না কেন তার থেকে যদি সম্ভাবনা তত্ত্বানুযায়ী পরস্পর নির্ভরতাশূন্যভাবে নমুনাসংগ্রহ করা হয়, তাহলে কয়েকটি সাধারণ সর্তসাপেক্ষে নমুনালব্ধ গড় বা সমষ্টির সম্ভাবনা-বিভাজন নমুনাসংখ্যাবৃদ্ধির সঙ্গে সঙ্গে নর্ম্যাল বিভাজনের প্রতি ক্রমাসন্ন (asymptotic) হয়। বৃহৎ নমুনাতত্ত্বে (large sample theory) এই ফলটি (result) কাজে লাগিয়ে প্রাক্কলন ও প্রকল্পবিচারে বহু সমাধান সম্ভব হয়েছে।

আবার অনেক সময় দেখা যায় যে, যদিও মূল চলটির বিভাজন নর্ম্যাল নয়, তবু তার বিশেষ কোন রূপান্তর (transformation) নেওয়া হলে সেই রূপান্তরিত চলটির বিভাজন অনেক ক্ষেত্রেই নর্ম্যাল হতে দেখা যায়। উদাহরণতঃ, অর্থশাস্ত্রে আলোচিত অনেক তথ্যকে ( যেমন আয় বা সম্পত্তিসূচক ) যদি $X$ চলের মান হিসেবে গণ্য করা হয়, তবে $\log X$-এর মানগুলি যে সারি উৎপন্ন করে তার বিভাজন প্রায়ই নর্ম্যাল প্রকৃতির হয়ে থাকে।

এই সমস্ত কারণে রাশিবিজ্ঞানে নর্ম্যাল বিভাজনের এত গুরুত্ব। বিজ্ঞানের বিভিন্ন শাখায় গৃহীত রাশিতথ্যের বিশ্লেষণে অনেক সময় নর্ম্যাল বিভাজনের গুণধর্মাবলী কাজে লাগিয়ে সমস্যা সমাধানের সার্থক প্রচেষ্টা দেখা যায়। শিল্প-ক্ষেত্রেও, যেমন গুণনিয়ন্ত্রণ ব্যাপারে নর্ম্যাল বিভাজনের বহুল প্রয়োগ দেখা যায়।

নর্ম্যাল বিভাজনের আলোচনার আরও বিশেষ সুবিধে হচ্ছে এর অতি সরল রূপপ্রকৃতি ও গুণধর্ম যাদের ভিত্তিতে অনেক গাণিতিক মান খুব সহজেই নির্ণয় করা যায় এবং প্রয়োজনীয় পারিসাংখ্যিক সারণী (statistical tables) ইত্যাদির সহজ সংকলন সম্ভব এবং বাস্তবিক এ ধরনের অনেক সারণী তৈরি হয়েছে।

অবশ্য এই আলোচনা থেকে এমন সিদ্ধান্ত করা ঠিক নয় যে, যে কোন পূর্ণকের বিভাজনই নর্ম্যাল হবে। বহু অ-নর্ম্যাল পূর্ণকের অস্তিত্ব রয়েছে এবং তাদের সম্পর্কে বিস্তারিত আলোচনা হয়েছে এবং আজও অনবরতই হয়ে চলেছে। কিন্তু এসব ক্ষেত্রেও অর্থাৎ যদি নিশ্চিত জানা থাকে যে, পূর্ণকটির বিভাজন নর্ম্যাল নয়, তবুও, নর্ম্যাল বিভাজনকে তাদের স্থূল বা প্রাথমিক আসন্ন

রূপ হিসেবে ধ'রে খানিকটা কাজ করা যায় এবং তাতে অনেকসময় খুব উল্লেখযোগ্য ভ্রান্তি সঞ্জাত হয় না। আবার অনেক পূর্ণক আছে যাদের বিভাজন এক-একটি সারি হিসেবে প্রকাশ করা যায় যা নর্ম্যাল সম্ভাবনা-ঘনত্ব-অপেক্ষক এবং অন্য কতগুলি অপেক্ষকের সংযোগে গঠিত। যেমন, গ্রাম ও শার্লিয়ারের সারি (Gram-Charlier Series), এজোয়ার্থের সারি (Edgeworth's Series), যাদের প্রতিটিই নর্ম্যাল সম্ভাবনা-ঘনত্ব-অপেক্ষক $f$ ও তার বিভিন্ন ক্রমের (order) অন্তর্কলকদের ঋজুরৈখিক যৌগসারি (linear compound series)।

### 8.3.6 পিয়ার্সনের রেখাবলী (Pearsonian System of Curves) :

এখন আমরা সাধারণভাবে কতগুলি অ-নর্ম্যাল বিভাজন সম্পর্কে আলোচনা করব। প্রকৃতপক্ষে আমরা একটি বিভাজন গোষ্ঠীর কথা বলব যার বিভিন্ন সদস্যের প্রত্যেকটিই এক-একটি পৃথক্ ধরনের পূর্ণক বিভাজন সূচিত করে এবং তাদের প্রতিটিকেই এক-একটি ভিন্ন প্রকৃতির রেখাদ্বারা রূপায়িত করা যায়। বাস্তবিক, এই সমস্ত বিভাজন রেখা মিলে একটি তথাকথিত পরিবার গ'ড়ে তুলেছে ব'লে মনে করা যায়। আমরা আরও দেখব যে, নর্ম্যাল রেখাও এই পরিবারভুক্ত একটি রেখা। এদেরকে কার্ল পিয়ার্সনের (Karl Pearson) বিভাজন-রেখাগোষ্ঠী (system of frequency curves) বলে অভিহিত করা যায়।

উনবিংশ শতাব্দীর শেষদিকে প্রখ্যাত জীববিজ্ঞানী কার্ল পিয়ার্সন চার্লস ডারউইনের বিবর্তনবাদের (Charles Darwin's Theory of Evolution) গাণিতিক ভিত্তিপ্রতিষ্ঠায় উদ্যোগী হয়ে অসংখ্য জীবপ্রজাতির (animal species) দেহের বিভিন্ন অবয়বের মাপজোখ নিয়ে তাঁর পরীক্ষাকাজ চালিয়েছিলেন। ঐ সব পরীক্ষা-নিরীক্ষায় একটি বিশেষ তথ্য উদ্ঘাটিত হয় যে একই জাতীয় জীবের একই অবয়বসংশ্লিষ্ট পরিমাপ সমুদয়ের ( যথা, মাথার খুলির ওজন, শিরদাঁড়ার দৈর্ঘ্য ইত্যাদি ) বিভাজন প্রায় সর্বদাই নর্ম্যাল বিভাজনের অনুরূপ। এর থেকে কার্ল পিয়ার্সন এই সিদ্ধান্তে পৌছান যে, কোন পূর্ণক যদি যথার্থ অন্তর্সম হয় তাহলে তার বিভাজন নর্ম্যাল হবে। এই অভিজ্ঞতাই তিনি স্বাভাবিক ব'লে মনে করেন এবং সেই থেকেই নর্ম্যাল বিভাজনের এরকম নামকরণ। কিন্তু প্রকৃতপক্ষে, তাঁর অভিজ্ঞতা যখন আরও ব্যাপক ও বিস্তৃততর হ'ল তখন তিনি তাঁর সিদ্ধান্ত পরিবর্তিত করতে বাধ্য হন, কারণ তিনি দেখতে

পান যে, সমীক্ষাসূত্রে প্রাপ্ত অনেক বিভাজনকে নর্ম্যাল বিভাজনের আওতার মধ্যে নিয়ে আসা যায় না। এজন্যে তাদের স্বরূপপ্রকৃতি আরও গভীরভাবে আলোচনা ক'রে তিনি দেখাতে সক্ষম হন যে, সেগুলোকে আরও ব্যাপকতর বিভাজন-গুচ্ছের কোন একটিকে দিয়ে রূপায়িত করা যায় এবং নর্ম্যাল বিভাজন হচ্ছে ঐ গুচ্ছেরই অন্তর্ভুক্ত একটি বিভাজন মাত্র। এখন আমরা ঐ বিভাজন-গুচ্ছের উৎপত্তি, তাদের স্বরূপ ও গুণধর্ম সংক্ষেপে আলোচনা করব।

বিভিন্ন প্রজাতির দেহাবয়বের মাপজোখ নিয়ে সমীক্ষার সূত্রে কার্ল পিয়ার্সন দেখতে পান যে, এদের মধ্যে অন্তর্সম তথ্যের ভিত্তিতে সংকলিত পরিসংখ্যা-বিভাজনগুলি এবং তৎসংশ্লিষ্ট পরিসংখ্যা-রেখাবলীর বৈশিষ্ট্য হচ্ছে এই যে, (1) তাদের একটি ক'রে মাত্র ভূয়িষ্ঠক বা সংখ্যাগরিষ্ঠমান রয়েছে এবং (2) তাদের ভুজের উভয় প্রান্তসীমায় উচ্চক্রম সংযোগ (high order contact) রয়েছে, যার ব্যবহারিক অর্থ হচ্ছে এই যে, পরিসংখ্যাগুলি প্রান্তীয় শ্রেণীঅন্তরগুলির জন্যে ধীরে ধীরে মসৃণতা বজায় রেখে কমতে থাকে এবং কখনই আকস্মিকভাবে ওঠা-নামা করে না। এর থেকে কার্ল পিয়ার্সন এই সিদ্ধান্তে উপনীত হন যে, এরূপ ধর্মবিশিষ্ট কোন নমুনালব্ধ বিভাজন যে পূর্ণক থেকে গৃহীত হবে তার বিভাজন এমন ধরনের হবে যে, তাকে এমন একটি সম্ভাবনা-ঘনত্ব-অপেক্ষক $f$ দ্বারা রূপায়িত করা যাবে যার জন্যে $\frac{df}{dx}$-এর মান 0 যখন $x=a$ এবং যখন $f(x)=0$. এখানে $a$ হচ্ছে $f$-এর একমাত্র ভূয়িষ্ঠক।

এর থেকে কার্ল পিয়ার্সন সিদ্ধান্তে আসেন যে, $f$-কে নিম্নলিখিত ধরনের একটি অন্তর্কলক সমন্বিত সমীকরণ সাহায্যে প্রকাশ করা যেতে পারে :

$$\frac{df}{dx}=\frac{(x-a)f}{F(x)}.$$

এখানে, $F(x)$ হচ্ছে যে কোন একটি অপেক্ষক। এরপর কিন্তু কার্ল পিয়ার্সন $F(x)$ সম্পর্কে একটি আপাত স্বেচ্ছাগৃহীত স্বীকরণের অবতারণা করেন। সেটি হচ্ছে এই যে, ম্যাক্লরীনের (Maclaurine) সারি অনুযায়ী $F(x)$-এর প্রসারণ (expansion) সম্ভব, অর্থাৎ $F(x)=F(0)+xF'(0)+\frac{x^2}{2!}F''(0)+\frac{x^3}{3!}F'''(0)+\cdots$ রূপে $F(x)$-এর প্রকাশন সম্ভব। তারপর তিনি আরও ধ'রে নেন যে, এই

প্রসারণে $x$ এর 2 এর অধিক সূচকসংশ্লিষ্ট পদগুলিকে বাদ দেওয়া যেতে পারে। এর ফলশ্রুতি হিসেবে লেখা যেতে পারে

$$\frac{df}{dx}=\frac{(x-a)f}{b_0+b_1x+b_2x^2}.$$

এর স্বপক্ষে যুক্তি দেখানো হয় যে,

(1) $F(x)$-এ আরও বেশী পদ রাখলে সমীকরণটির সমাধান জটিল হবে অথচ এই পর্য্যন্ত মাত্র পদ রাখলে সমাধান সহজ হবে এবং ঐ সহজ সমাধান সাহায্যে যে সমস্ত সম্ভাবনা-ঘনত্ব-অপেক্ষক পাওয়া যাবে তাদের মাধ্যমেই সাধারণভাবে প্রত্যাশিত সব পরিসংখ্যা-বিভাজনকে রূপায়িত করা যাবে এবং (2) অধিক পদ রাখলে সমীকরণটিতে আরও অধিক পূর্ণকাঙ্কের আবির্ভাব হবে এবং তাদের প্রাক্কলনে অধিকতর ক্রমের পরিঘাত ব্যবহার করতে হবে এবং তাতে নমুনাগত ভ্রান্তি অত্যন্ত বেশী হয়ে পড়বে।

পিয়ার্সনের অন্তর্কলকযুক্ত সমীকরণ $\frac{df}{dx}=\frac{(x-a)f}{b_0+b_1x+b_2x^2}$-এর সমাধান সহজেই নির্ণয় করা যায়। কারণ, লেখা যায় যে,

$$\frac{df}{f}=\frac{(x-a)\,dx}{b_0+b_1x+b_2x^2},$$

$$\int\frac{df}{f}=\int\frac{(x-a)\,dx}{(b_0+b_1x+b_2x^2)}+k \qquad [k \text{ হচ্ছে সমাকলন-জনিত ধ্রুবক}]$$

$$\text{বা}\ \log f=\int\frac{(x-a)\,dx}{(b_0+b_1x+b_2x^2)}+k$$

$$\text{বা}\quad f=c.\exp\left[\int\frac{(x-a)\,dx}{(b_0+b_1x+b_1x^2)}\right], \quad [\log_e c=k]$$

এটিই হচ্ছে সমীকরণটির সাধারণ সমাধান। বিশেষতর সমাধান পেতে হলে $(b_0+b_1x+b_2x^2)$ এই দ্বিঘাত প্রকাশনটির (quadratic expression) মূলদ্বয়ের স্বরূপের ওপর নির্ভর করতে হবে। বাস্তবিক ঐ মূলদ্বয়ের স্বরূপের ভিন্নতা অনুযায়ী বিভিন্ন পিয়ার্সনীয় বিভাজন-রেখার উৎপত্তি হবে।

এই আলোচনা বিস্তৃততর করতে হলে পিয়ার্সনের নিরিখ (criterion) $\kappa=\frac{b_1^2}{4b_0b_2}$-এর বিভিন্ন মান বিবেচনা করতে হবে।

(1) যদি $\kappa<0$ হয়, অর্থাৎ যদি $b_0+b_1x+b_2x^2$-এর মূলদ্বয় প্রকৃত,

অসমান ও বিপরীত চিহ্নবিশিষ্ট হয় অর্থাৎ যদি $b_0$ ও $b_2$ বিপরীত চিহ্নবিশিষ্ট হয়,তাহলে যে সমস্ত রেখা পাওয়া যায় তাদের বলা হয় পিয়ার্সনের প্রথম প্রকার রেখা (Type I curve)।

(2) যদি $\kappa > 1$ হয় অর্থাৎ যদি $b_0+b_1x+b_2x^2$-এর মূল-দুটি প্রকৃত (real) ও সমচিহ্নবিশিষ্ট হয় [ এক্ষেত্রে অবশ্যই $b_0$ ও $b_2$ উভয়েই ঋণাত্মক বা উভয়েই ধনাত্মক হবে ], তখন যে রেখা পাওয়া যায়, তাকে বলে পিয়ার্সনের ষষ্ঠ প্রকার রেখা (Type VI curve)।

(3) যদি $0 < \kappa < 1$ হয় অর্থাৎ যদি $b_0+b_1x+b_2x^2$-এর মূল-দুটি কল্পিত (imaginary) বা মিশ্ররাশি (complex) হয়, তাহলে যে রেখা পাওয়া যায়, তাকে বলে চতুর্থ প্রকার রেখা (Type IV curve)।

এই তিন শ্রেণীর রেখাকেই মুখ্যশ্রেণীর রেখা (curves of main type) বলে। এছাড়া আরও কয়েকটি গৌণশ্রেণীর রেখা (transition type of curves) পাওয়া যায়।

(4) যদি $b_2=0$ অর্থাৎ $\kappa=\pm\infty$ হয়, তাহলে যে রেখা পাওয়া যায়, তাকে বলে তৃতীয় প্রকার রেখা (Type III curve)।

(5) যদি $\kappa=1$ হয়, তবে যে রেখা পাওয়া যায়, তাকে বলে পঞ্চম প্রকার রেখা (Type V curve)। এক্ষেত্রে $b_0+b_1x+b_2x^2$-এ এর মূলদ্বয় সমমানযুক্ত।

(6) যদি $b_1=0$ এবং $b_0$ ও $b_2$ বিপরীত বা সমচিহ্নবিশিষ্ট হয়, তবে যথাক্রমে দ্বিতীয় ও সপ্তম প্রকার রেখা ( Type II and Type VII curves) পাওয়া যায়।

সবশেষে, (7) যদি $b_1=0=b_2$ হয়, তাহলে নর্ম্যাল রেখা পাওয়া যায়, কারণ এই ক্ষেত্রে

$$\frac{df}{dx}=\frac{xf}{b_0}$$ [ কারণ দেখানো যায় যে, সব রকম রেখার জন্যে $b_1=a$]

ফলে, $f=c.\exp\left[\frac{1}{b_0}\int x\,dx\right]=c.\exp\left(\frac{x^2}{2b_0}\right)$.

স্পষ্টতঃ এটিই হচ্ছে পূর্বালোচিত নর্ম্যাল-ঘনত্ব-অপেক্ষকের রূপ। পিয়ার্সনের অন্তর্কলকযুক্ত সমীকরণটির সমাধান নির্ণয় ক'রে ওপরে যেমন বলা হয়েছে তেমনি বিভিন্ন পরিস্থিতিতে যে বিভিন্ন প্রকার বিভাজন-রেখা (frequency curves) পাওয়া যায়, সেগুলির সমীকরণ হচ্ছে নিম্নবর্ণিতরূপ :

### 8.3.6.1 বিভিন্ন পিয়ার্সনীয় রেখার সমীকরণ:

**প্রথম প্রকার রেখা**

$$f(X)=f_0\left(1+\frac{X}{a_1}\right)^{m_1}\left(1-\frac{X}{a_2}\right)^{m_2};\ -a_1 \leqslant X \leqslant a_2.$$

এখানে $\dfrac{m_1}{a_1}=\dfrac{m_2}{a_2}=\dfrac{m_1+m_2}{a_1+a_2}$, এবং $X=x-a$

অর্থাৎ $X$-কে ভূয়িষ্ঠক $a$ থেকে মাপা হয়েছে। পরিভাষানুযায়ী $X$-এর মূল (origin) হচ্ছে ভূয়িষ্ঠক $a$-তে। এটি একটি অপ্রতিসম (asymmetrical) রেখা এবং এর আকৃতি ঘণ্টার মতো (bell-shaped) যদি $m_1$ ও $m_2$ উভয়েই ধনাত্মক হয়। $m_1$ ও $m_2$ উভয়েই ঋণাত্মক হলে এর আকৃতি U-এর মতো। যদি $m_1>0$ ও $m_2<0$ হয়, তবে এর আকৃতি (ɾ) উল্টো $J$-এর মতো এবং যদি $m_1<0$ ও $m_2>0$ হয়, তবে এর আকৃতি $J$-এর মতো।

**ষষ্ঠ প্রকার রেখা**

$f(X)=f_0\,(X-a)^{q_2}X^{-q_1},\ a\leqslant X<\infty$; $X$-এর মূল হচ্ছে রেখাটি যেখান থেকে সুরু হয়েছে (start of the curve) তার থেকে $a$ একক পূর্বে। এটিও অপ্রতিসম ও এর আকৃতি ঘণ্টার ন্যায় যদি $q_2>0$ হয়। কিন্তু $q_2<0$ হলে এটি $J$ আকৃতিবিশিষ্ট হবে।

**চতুর্থ প্রকার রেখা**

$$f(X)=f_0\left(1+\frac{X^2}{a^2}\right)^{-m}e^{-\nu\tan^{-1}\frac{x}{a}},\ -\infty<X<\infty.$$

$X$-এর মূল হচ্ছে গড় বিন্দু থেকে $\dfrac{\nu a}{2m-2}$ একক উর্ধ্বে। এটি অপ্রতিসম এবং সর্বদাই ঘণ্টাকৃতিবিশিষ্ট।

**তৃতীয় প্রকার রেখা**

$$f(X)=f_0\,e^{-\gamma X}\left(1+\frac{X}{a}\right)^{\gamma X}\ -a\leqslant X<\infty$$

$x$-এর মূল হচ্ছে ভূয়িষ্ঠকে। $\gamma a=p$-এর ধনাত্মক মানের জন্যে এই রেখার আকৃতি ঘণ্টার ন্যায় এবং $p$-এর ঋণাত্মক মানের জন্যে এটি $J$-আকৃতিবিশিষ্ট। এটিও অপ্রতিসম।

**পঞ্চম প্রকার রেখা**

$$f(X)=f_0\, X^{-p}\, e^{-\frac{\gamma}{X}},\ 0 < X < \infty.$$

$X$-এর মূল রেখাটির সুরুতে। এই রেখাটি সর্বদাই ঘণ্টাকৃতিবিশিষ্ট। এটিও অপ্রতিসম।

**দ্বিতীয় প্রকার রেখা**

$$f(X)=f_0\left(1-\frac{X^2}{a^2}\right)^m,\ -a < X < a.$$

এর মূল হচ্ছে গড় বিন্দুতে। এটি গড় বিন্দু $O$-এর উভয় পার্শ্বে প্রতিসম এবং $m$-এর ধনাত্মক ও ঋণাত্মক মানের জন্যে রেখাটির আকৃতি যথাক্রমে ঘণ্টা এবং U-এর মতো।

**সপ্তম প্রকার রেখা**

$$f(X)=f_0\left(1+\frac{X^2}{a^2}\right)^{-m},\ -\infty < X < \infty.$$

এর মূল হচ্ছে গড় বিন্দুতে অর্থাৎ $O$-তে। এটিও মূলবিন্দু $O$-এর উভয়পার্শ্বে প্রতিসম এবং এটি সর্বদাই ঘণ্টাকৃতিবিশিষ্ট।

নর্ম্যালরেখার রূপপ্রকৃতি সম্পর্কে আগেই বিস্তৃত আলোচনা করা হয়েছে।

আমরা দেখেছি যে, অন্তর্কলকযুক্ত সমীকরণ (differential equation) $\frac{df}{dx}=\frac{(x-a)f}{b_0+b_1x+b_2x^2}$ থেকে নির্ণেয় পরিসংখ্যা রেখা (frequency curve) $f$-এর স্বরূপ $\kappa=\frac{b_1^2}{4b_0b_2}$ এর বিভিন্ন মানের দ্বারা নির্ধারিত হয়। এখানে $x$-এর গড় শূন্য ব'লে ধ'রে নেওয়া যেতে পারে ( প্রয়োজন হলে মাপনা মূলবিন্দু পরিবর্তিত ক'রে নিয়ে ) অর্থাৎ আমরা স্বীকার করি যে,

$$\int_{-\infty}^{\infty} x f(x)\, dx = 0.$$

কাজেই $\mu_r=\int_{-\infty}^{\infty} x^r f(x)\, dx$ হচ্ছে $f(x)$-এর গড় কেন্দ্রিক $r$-তম পরিঘাত।

এখন, দেখানো যায় যে, $a, b_0, b_1, b_2$ এই চারটি ধ্রুবককে $\mu_2, \mu_3$ ও $\mu_4$-এর মাধ্যমে প্রকাশ করা সম্ভব এবং $a=b_1$. বাস্তবিক, দেখানো যায় যে,

$$b_0=\frac{-(4\beta_2-3\beta_1)\mu_2}{2(5\beta_2-6\beta_1-9)},\ a=b_1=\frac{-\sigma\sqrt{\beta_1}(\beta_2+3)}{2(5\beta_2-6\beta_1-9)},\ \sigma=+\sqrt{\mu_2}$$

এবং $b_2 = \dfrac{-(2\beta_2 - 3\beta_1 - 6)}{2(5\beta_2 - 6\beta_1 - 9)}$.

কাজেই $\kappa = \dfrac{b_1^2}{4b_0 b_2} = \dfrac{\beta_1(\beta_2+3)^2}{4(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)}$.

এখন, আমরা জানি যে, পিয়ার্সনের রেখার ধরন হবে যথাক্রমে

প্রথম প্রকার, যখন $\kappa < 0$ অর্থাৎ $2\beta_2 - 3\beta_1 - 6 < 0$;

চতুর্থ প্রকার, যখন $0 < \kappa < 1$ অর্থাৎ $(2\beta_2 - 3\beta_1 - 6) > 0$ এবং
$\beta_1(\beta_2+3)^2 < 4(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)$;

ষষ্ঠ প্রকার, " $\kappa > 1$ অর্থাৎ $\beta_1(\beta_2+3)^2 > 4(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)$;

পঞ্চম প্রকার, " $\kappa = 1$ " $\beta_1(\beta_2+3)^2 = 4(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)$;

তৃতীয় প্রকার, " $b_2 = 0, \kappa = +\infty$, অর্থাৎ $2\beta_2 - 3\beta_1 - 6 = 0$;

দ্বিতীয় প্রকার, " $b_1 = 0$ এবং $b_0$ ও $b_2$ বিপরীত চিহ্নযুক্ত অর্থাৎ
যখন $\beta_1 = 0, \beta_2 < 3$;

সপ্তম প্রকার, " $b_1 = 0$ এবং $b_0$ ও $b_2$ সমচিহ্নযুক্ত অর্থাৎ
যখন $\beta_1 = 0, \beta_2 > 3$;

এবং নর্ম্যাল বা গাউসীয়, যখন $b_1 = 0$ ও $b_2 = 0$ অর্থাৎ $\beta_1 = 0$ ও $\beta_2 = 3$. নর্ম্যাল রেখার চর্চায় মনীষী গাউসের (Gauss) যথেষ্ট অবদান রয়েছে এবং এক সময় এর সঙ্গে তাঁর নাম অঙ্গাঙ্গিভাবে জড়িত ছিল। সেজন্যে একে অনেক সময় গাউসীয় রেখাও (Gaussian curve) বলা হয়ে থাকে।

কাজেই দেখা যাচ্ছে যে, পিয়ার্সনের রেখা পরিবারভুক্ত বিভিন্ন সদস্য $\beta_1$ ও $\beta_2$-এর পারস্পরিক সম্পর্কের সূত্রে নির্দিষ্ট হয়। এই তথ্যকে নিম্নবর্ণিতভাবে কাজে লাগানো হয়ে থাকে। $\beta_1$ ও $\beta_2$-কে যথাক্রমে ভুজ ও কোটি ধ'রে একটি লেখচিত্র আঁকা হলে তা বিভিন্ন অঞ্চলে বিভক্ত ব'লে মনে করা যেতে পারে, যাদের মধ্যে পৃথক্ পৃথক্ ভাবে উল্লিখিত $\beta_1$ ও $\beta_2$-সংশ্লিষ্ট বিভিন্ন সম্পর্কগুলি খাটবে। কাজেই ঐ এক-একটি অঞ্চলকে পিয়ার্সনের বিভিন্ন ধরনের রেখাভুক্ত অঞ্চল ব'লে ধরা যায়। এই লেখকে বলে $(\beta_1$-$\beta_2)$ চিত্র। হার্টলে ও পিয়ার্সনের (Hartley and Pearson) সংকলিত Biometrika Table-এ এই চিত্র আঁকা আছে। এখন, যদি আমাদের হাতে কোন পরিসংখ্যা-বিভাজন থাকে, তবে তার থেকে $\beta_1$ ও

$\beta_2$ অঙ্কের নমুনাগত প্রাক্-কলক $\hat{\beta}_1 = \frac{m_3}{m_2^3}$, $\hat{\beta}_2 = \frac{m_4}{m_2^2}$ [ এখানে $m_r$ হচ্ছে নমুনালব্ধ $r$-তম গড়কেন্দ্রিক পরিঘাত ] এর মান কষে বের ক'রে Biometrika Table-এ অঙ্কিত ($\beta_1$-$\beta_2$) চিত্রে ($\hat{\beta}_1$, $\hat{\beta}_2$) বিন্দুটি ঐ চিত্রের কোন্ অঞ্চলে পড়ছে তা দেখে জানা যায় ঐ নমুনাবিভাজনটি যে অজ্ঞাত পূর্ণক থেকে নেওয়া হয়েছে সেই পূর্ণকের বিভাজনটিকে পিয়ার্সনের রেখামালার কোন্‌টি দিয়ে সূচিত করা সমীচীন হবে। এই হচ্ছে ($\beta_1$-$\beta_2$) চিত্রের প্রয়োগভিত্তিক উপযোগিতা। এখানে অবশ্য একটি কথা মনে রাখা দরকার যে, ($\beta_1$-$\beta_2$) চিত্রের ($\beta_1$, $\beta_2$) বিন্দুগুলির $\beta_1$ ও $\beta_2$ হচ্ছে এক একটি পূর্ণক-সংশ্লিষ্ট অঙ্ক। কিন্তু নমুনালব্ধ $\hat{\beta}_1$, $\hat{\beta}_2$ হচ্ছে যথাক্রমে আসল $\beta_1$ ও $\beta_2$-এর প্রাক্-কলক মাত্র। কাজেই $\hat{\beta}_1$ ও $\hat{\beta}_2$-এর মধ্যে নমুনাগত ভ্রান্তি থাকবে। ফলে, নমুনাসঞ্জাত ($\hat{\beta}_1$, $\hat{\beta}_2$) বিন্দুটি ($\beta_1$-$\beta_2$) চিত্রের একটি বিশেষ অঞ্চলে ( উদাহরণতঃ তৃতীয় প্রকার রেখার জন্যে নির্দিষ্ট অঞ্চলে ) পড়লেও যে পূর্ণক থেকে ঐ নমুনাটি এসেছে তা ঐ বিশেষ অঞ্চলের জন্য নির্দিষ্ট ধরনের পিয়ার্সনীয় রেখা দ্বারা নির্দেশযোগ্য নাও হতে পারে। এই জন্যেই আবার এটা বলা যাবে যে, যদি ($\hat{\beta}_1$, $\hat{\beta}_2$) কোন একটি বিশেষ অঞ্চলের ( ধরা যাক, দ্বিতীয় প্রকার রেখার অঞ্চল ) মধ্যে না পড়ে যদি তার কাছাকাছি পড়ে তাহলেও নমুনাভ্রান্তির কথা মনে রেখে যে পূর্ণক থেকে নমুনাটি এসেছে তাকে ঐ বিশেষ অঞ্চলের জন্যে নির্দিষ্ট পিয়ার্সন রেখা দিয়ে নির্দিষ্ট করার চেষ্টা করা যেতে পারে। তাহলে সবশেষে সাযুজ্য-নিরূপণ-পদ্ধতি প্রয়োগ ক'রে দেখতে হবে আসলে কোন পিয়ার্সন রেখা দিয়ে পূর্ণকটিকে সঙ্গতভাবে সূচিত করা যায় কিনা।

## 8.3.7 উদাহরণমালা :

এখন, আমরা প্রদত্ত পরিসংখ্যা বিভাজনের সঙ্গে বাইনোমিয়াল তত্ত্বগত বিভাজনের সাযুজ্য-নিরূপণ-পদ্ধতির প্রয়োগ উদাহরণ সাহায্যে আলোচনা করব।

**উদা. 8.1** একটি ফুলকপি ক্ষেতকে 80টি সমান্তরাল সারিতে বিভক্ত ক'রে তার প্রত্যেকটিকে আবার ছোট ছোট 10টি ক'রে খণ্ডে ভাগ ক'রে তার প্রত্যেকটিতে ফুলকপির বীজ পুঁতে 7 দিন পরে তাদের মধ্যে কতগুলিতে অঙ্কুরোদ্গম হয়েছে দেখতে গিয়ে নিম্নবর্ণিত বিভাজনটি দেখা গেল।

সারণী 8.1

| সারিতে অঙ্কুরিত বীজের সংখ্যা | সারির পরিসংখ্যা |
|---|---|
| $x$ | $f_x$ |
| 0 | 6 |
| 1 | 20 |
| 2 | 28 |
| 3 | 12 |
| 4 | 8 |
| 5 | 6 |
| 6 বা ততোধিক | 0 |

এখানে আমরা বলতে পারি যে, প্রতি সারিতে 10টি ক'রে বীজ মাটিতে পুঁতে বেরণুলীয় পরীক্ষা চালানো হয়েছে যার প্রতিটি ফলাফল হচ্ছে দুটি বিকল্পের মধ্যে একটি এবং বিকল্পরূপ-দুটি হচ্ছে বীজের অঙ্কুর উদ্গত হওয়া বা না হওয়া। এখানে বীজ অঙ্কুরিত হওয়াকে সার্থকতা এবং তার অন্যথাকে ব্যর্থতা বলা যেতে পারে। এখানে এরকম 80টি বিভিন্ন বেরণুলীয় পরীক্ষা চালানো হয়েছে ব'লে ধরা যায়। এখন একটি তত্ত্বগত বাইনোমিয়াল বিভাজনের সঙ্গে প্রদত্ত বিভাজনটির সাযুজ্য নির্ধারণ করতে হলে নিম্নবর্ণিতভাবে অগ্রসর হওয়া যায়।

এখানে 80টি পরীক্ষণের প্রতিটিতে প্রচেষ্টার সংখ্যা $n=10$.

নমুনাগড় $\bar{x}=\dfrac{\Sigma x f_x}{\Sigma f_x}=2{\cdot}175,\quad \hat{p}=\dfrac{\bar{x}}{n}={\cdot}2175,\ \hat{q}=1-\hat{p}={\cdot}7825,$

মোট পরিসংখ্যা $N=\Sigma f_x=80.$

$$\hat{f}(x)=f(x\,;\ \hat{p})=\binom{n}{x}\hat{p}^x(1-\hat{p})^{n-x}$$

$$=\binom{n}{x}g(x\,;\,\hat{p})\quad (\text{ধর})$$

এখানে $g(x\,;\,\hat{p})=\hat{p}^x(1-\hat{p})^{n-x}$;

তাহলে $\log g(x\,;\,\hat{p})=x\log\hat{p}+(n-x)\log(1-\hat{p}).$

## সারণী 8.2

### তত্ত্বগত বাইনোমিয়াল বিভাজনের সঙ্গে প্রদত্ত পরিসংখ্যা-বিভাজনের সাযুজ্য নিরূপণ

| $x$ | $\binom{n}{x}$ | $x \log \hat{p}$ | $(n-x)$ $\log (1-\hat{p})$ | $x \log \hat{p}$ $+(n-x)$ $\log (1-\hat{p})$ | লব্ধপংক্তি (5)-এর প্রতি-লগারিদম | $f(x\,;\,\hat{p})$ | প্রত্যাশিত পরিসংখ্যা | অবেক্ষিত পরিসংখ্যা |
|---|---|---|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| 0 | 1 | 0 | −1·06516 | −1·06516 | ·08607 | ·08607 | 6·89 | 6 |
| 1 | 10 | −·66254 | −·95864 | −1·62118 | ·02387 | ·23868 | 19·09 | 20 |
| 2 | 45 | −1·32508 | −·85213 | −2·17721 | ·00665 | ·29925 | 23·94 | 28 |
| 3 | 120 | −1·98762 | −·74561 | −2·73323 | ·00185 | ·22179 | 17·74 | 12 |
| 4 | 210 | −2·65016 | −·63909 | −3·28926 | ·00051 | ·10710 | 8·57 | 8 |
| 5 | 252 | −3·31270 | −·53258 | −3·84528 | ·00014 | ·03528 | 2·82 | 6 |
| 6 বা ততোধিক | — | — | — | — | — | ·01183* | ·95 | 0 |

$$* \sum_{x=6}^{10} f(x\,;\,\hat{p}) = 1 - \sum_{x=0}^{5} f(x\,;\,\hat{p})$$

এই সাযুজ্য-নিরূপণ সহজতরভাবে 8.3 সারণীর সাহায্যে করা যায়।

এখন একটি প্রদত্ত পরিসংখ্যা-বিভাজনের সঙ্গে পোয়াসঁ বিভাজনের সাযুজ্য নিরূপণ কী ভাবে করা যায় উদাহরণ সাহায্যে দেখা যাক।

**উদা. 8.2**　7·5 সেকেণ্ড স্থায়ী এক একটি কালবিরতিতে একটি তেজস্ক্রিয় বস্তুর (radioactive particle) উপাদানসমূহের কতগুলি একটি নির্দিষ্ট স্থানে পৌঁছায় তা দেখার জন্যে একটি পরীক্ষাকার্য মোট 2610 বার চালিয়ে 8.4 সারণীতে প্রদত্ত পরিসংখ্যা-বিভাজনটি পাওয়া যায়।

সারণী 8.3

| $x$ | $\frac{n-x+1}{x}$ | $\frac{n-x+1}{x} \cdot \frac{\hat{p}}{\hat{q}}$ | $\hat{f}(x) = (3) \times f(x-1)$ | প্রত্যাশিত পরিসংখ্যা $= N\hat{f}(x)$ $= Nx(4)$ | অবেক্ষিত পরিসংখ্যা |
|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) | (6) |
| 0 | — | — | ·086* | 6·89 | 6 |
| 1 | 10 | 2·78 | ·239 | 19·12 | 20 |
| 2 | 4·50 | 1·25 | ·299 | 23·92 | 28 |
| 3 | 2·67 | ·74 | ·221 | 17·68 | 12 |
| 4 | 1·75 | ·49 | ·108 | 8·64 | 8 |
| 5 | 1·20 | ·33 | ·036 | 2·88 | 6 |
| 6 বা ততোধিক | — | — | ·011** | ·88 | 0 |

$$*f(0\,;\,\hat{p}) = \hat{q}^{10} = {\cdot}086$$

$$**\sum_{x=6}^{10} f(x\,;\,\hat{p}) = 1 - \sum_{x=0}^{5} \hat{f}(x)$$

সারণী 8.4

| নির্দিষ্টস্থানে উপস্থিত তেজস্ক্রিয় বস্তুর কণিকাপুঞ্জের সংখ্যা ($K$) | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | $\geqslant 10$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| পরিসংখ্যা ($N_k$) | 57 | 203 | 385 | 525 | 532 | 408 | 273 | 139 | 45 | 27 | 6 |

এখন, 7·5 সেকেণ্ডের এক একটি সময়-স্থায়িত্বকে আরও ক্ষুদ্র ক্ষুদ্র অসংখ্য কালবিরতির এক একটি গুচ্ছ হিসেবে দেখা যেতে পারে, যাতে সেই ক্ষুদ্রতর সময়দৈর্ঘ্যগুলি এত ছোট যে তার এক একটিতে সর্বাধিক একটি বস্তুকণিকা নির্দিষ্ট স্থানে গিয়ে পৌঁছাতে পারে। তাহলে এই এক একটি ক্ষুদ্র কালাংশকে

এক একটি পরীক্ষণ প্রচেষ্টা মনে করতে পারি যাতে যদি তার মধ্যে একটি বস্তু-কণিকা নির্দিষ্ট স্থানে পৌছায় তাহলে প্রচেষ্টাটি সার্থক ও অন্যথায় সেটি ব্যর্থ হ'ল ব'লে স্বীকার করা যায়। এখানে পরীক্ষণ প্রচেষ্টাগুলিকে সম্ভাবনা তত্ত্বানুযায়ী পরস্পর অনধীন ব'লে মনে করা হবে। তাহলে পরীক্ষণসংখ্যা $n$ অসীম হবে এবং প্রতি প্রচেষ্টায় সার্থকতার সম্ভাবনা $p$ খুবই অল্প হবে অথচ $np$ বিশেষ বড় বা ছোট হবে না, কারণ $np$-এর একটি মোটামুটি হিসেব হবে

$$\frac{\sum kN_k}{\sum N_k} = \frac{10094}{2610} = 3{\cdot}867.$$

কাজেই এটা আশা করা অন্যায় হবেনা যে, একটি পোয়াসঁ বিভাজনের সঙ্গে এই প্রকৃত বিভাজনটির একটি সাযুজ্য থাকবে। এখন এই চেষ্টা করতে গিয়ে 8.5 সারণীটি গঠন করা যাক। এই উদ্দেশ্যে পোয়াসঁ সম্ভাবনা-ভর-অপেক্ষক $f(x) = f(x\ ;\ m)$ এর পূর্ণকাঙ্ক $m$-এর প্রাক্-কলক হিসেবে

$$m = \bar{x} = \frac{\Sigma\, kN_k}{\Sigma\, N_k} = 3{\cdot}867\text{-কে নেওয়া হবে}$$

8.5 সারণীতে (4) ও (5) নম্বর স্তম্ভ-দুটি তুলনা করলে দেখা যায় যে এদের মধ্যে মোটামুটি ভালো মিল রয়েছে এবং এটাই প্রত্যাশিত।

বাস্তবিক, যে সমস্ত পরিসংখ্যা-বিভাজন খুব কদাচিৎ দৃষ্ট ঘটনার সঙ্গে জড়িত তাদের সঙ্গেই পোয়াসঁ বিভাজনের সাযুজ্য লক্ষ্য করা যায়। দু'একটি উদাহরণ উল্লেখ করা যাক।

কোন শহরের একটি জনবহুল রাস্তায় কোন একটি বিশেষ মাসে সংঘটিত দৈনিক মারাত্মক দুর্ঘটনার পরিসংখ্যা-বিভাজন যদি বিবেচনা করা যায় তবে তার সঙ্গে পোয়াসঁ বিভাজনের সাযুজ্য থাকা খুবই সঙ্গত। এখানে প্রত্যেক দিন (24 ঘণ্টা)-কে অসংখ্য ক্ষুদ্র ক্ষুদ্র দৈর্ঘ্যবিশিষ্ট কালাংশে বিভক্ত মনে করা যেতে পারে যার প্রতিটি মুহূর্তের স্থায়িত্বকাল এত স্বল্প যে তার মধ্যে ঐ রাস্তায় হয় সর্বাধিক একটি মারাত্মক দুর্ঘটনা ঘটবে আর নয় ত একটিও ঘটবে না। তাহলে আমরা মনে করতে পারি যে, এখানে প্রতিদিন অসংখ্য বেরণুলীয় পরীক্ষণ প্রচেষ্টা চালানো হ'ল যার প্রতিটি প্রচেষ্টা হচ্ছে এক একটি কালাংশ এবং প্রতিটি প্রচেষ্টা সার্থকতায় পর্যবসিত হয় যদি কোন কালাংশে একটি মারাত্মক দুর্ঘটনা ঘটে এবং সেটি ব্যর্থ হয় যদি ঐ সময়ে কোন মারাত্মক দুর্ঘটনা সংঘটিত না হয়।

### সারণী 8.5
### একটি প্রদত্ত বিভাজনের সঙ্গে একটি তত্ত্বগত পোয়াসঁ বিভাজনের সাযুজ্য নিরূপণ

| $x$ | $\frac{\hat{m}}{x}$ | $f(x)=\frac{\hat{m}}{x}f(x-1)$ | প্রত্যাশিত পরিসংখ্যা | নমুনা পরিসংখ্যা |
|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) |
| 0 | – | ˙0209 | 54˙55 | 57 |
| 1 | 3˙867 | ˙0808 | 211˙00 | 203 |
| 2 | 1˙934 | ˙1563 | 408˙00 | 385 |
| 3 | 1˙289 | ˙2015 | 525˙94 | 525 |
| 4 | ˙967 | ˙1948 | 508˙51 | 532 |
| 5 | ˙773 | ˙1507 | 393˙30 | 408 |
| 6 | ˙645 | ˙0971 | 253˙51 | 273 |
| 7 | ˙552 | ˙0537 | 140˙05 | 139 |
| 8 | ˙483 | ˙0259 | 67˙70 | 45 |
| 9 | ˙430 | ˙0111 | 31˙29 | 27 |
| $\geq$10 | – | ˙0070* | 18˙26 | 16 |

$$^{k}\sum f(x)=1-\sum f(x)$$

আবার ঐ প্রচেষ্টাগুলিকে পরস্পর নির্ভরতাশূন্য ব'লেও ধরা যেতে পারে, কারণ কালাংশগুলি অবিচ্ছিন্নভাবে পার হয়ে যাচ্ছে এবং তার কোনটিতে দুর্ঘটনা হওয়া না হওয়া অন্য কালাংশে দুর্ঘটনার সংঘটনকে প্রভাবিত করবে না ব'লে আশা করা যায় [ অবশ্য এটা ভাবা আশ্চর্য্য নয় যে, একবার ঐ রাস্তায় দুর্ঘটনা ঘটলে অন্ততঃ কিছুক্ষণ সেখানে গাড়ীচালক, পথচারী ইত্যাদি সকলেই একটু বেশী সতর্ক হয়ে দুর্ঘটনায় সম্ভাবনাকে কমিয়ে দিতে পারেন ; কিন্তু আমরা ধ'রে নিচ্ছি যে, রাস্তাটি যথেষ্ট দীর্ঘ এবং অত্যন্ত জনবহুল ও গাড়ী-ঘোড়া অত্যন্ত বেশী চলে, যার ফলে রাস্তার কোন এক অঞ্চলে দুর্ঘটনা ঘটলে সঙ্গে সঙ্গে সে সংবাদ রাস্তার সব অংশে ছড়িয়ে পড়ে না ও ফলে অতি সতর্কতার ভাব সৃষ্টি হয় না ও দুর্ঘটনার

সম্ভাবনার লক্ষণীয় কোন হ্রাসবৃদ্ধি হয় না ]। তাহলে আমরা মনে করতে পারি যে, এখানে বেরণুলীয় প্রচেষ্টাগুলি সংখ্যায় খুব বেশী এবং প্রতি প্রচেষ্টায় সার্থকতার সম্ভাবনা খুব কম অথচ ঐ মাসে দৈনিক গড় মারাত্মক দুর্ঘটনার সংখ্যা বেশী নয়। ফলে দৈনিক মারাত্মক দুর্ঘটনার সংখ্যা নামক যে চলটির কথা আমরা ভাবছি, তা পোয়াসঁ বিভাজন অনুসরণ করবে ব'লে আমরা সঙ্গতভাবেই ভাবতে পারি। এখন $n$ ও $p$ যথাক্রমে যদি প্রচেষ্টাসংখ্যা ও প্রতি প্রচেষ্টায় সার্থকতার সম্ভাবনা বোঝায় ( যাকে আমরা প্রতি প্রচেষ্টার জন্যে ধ্রুবক ব'লে স্বীকার ক'রে নিচ্ছি ) তাহলে $np = m$ এর প্রাক্-কলক হিসেবে নেব ঐ মাসের জন্যে গড় দৈনিক মারাত্মক দুর্ঘটনার সংখ্যা এবং ঐ সংখ্যা সাধারণতঃ খুব বেশী হয় না। কাজেই এক্ষেত্রে পোয়াসঁ বিভাজন পরিলক্ষিত পরিসংখ্যা-বিভাজনের সঙ্গে ভালো সাযুজ্য রক্ষা করবে ব'লে আশা করা যায়।

দ্বিতীয় উদাহরণস্বরূপ ধরা যাক যে, 50 পৃষ্ঠার একটি বই আছে যার প্রতি পৃষ্ঠায় অনেকগুলি ক'রে শব্দ ছাপা রয়েছে। এখন মুদ্রাকর যদি যথেষ্ট সতর্ক প্রকৃতির এবং আপন কাজে সুদক্ষ হন তবে মুদ্রণ-প্রমাদ খুব অল্পই ঘটবে। কিন্তু তৎসত্ত্বেও বইতে কিছু কিছু ছাপার ভুল থাকা স্বাভাবিক যদিও সেটি একটি বিরল ঘটনা বলেই স্বীকার্য। এখন ঐ বইয়ের প্রতি পাতায় মুদ্রণ-প্রমাদের পরিসংখ্যা-বিভাজন যদি নির্ণয় করা হয় তাহলে তার সঙ্গে একটি পোয়াসঁ বিভাজনের সাযুজ্য থাকা স্বাভাবিক। এখানে কোন একটি পৃষ্ঠায় যতগুলি অক্ষর আছে ঐ পৃষ্ঠার জন্যে ততগুলি প্রচেষ্টা আছে ব'লে মনে করা যেতে পারে এবং কোন অক্ষর যদি মুদ্রণ-প্রমাদদুষ্ট হয় তাহলেই বলব যে সংশ্লিষ্ট প্রচেষ্টাটি সার্থক হ'ল এবং কোন অক্ষর নির্ভুলভাবে মুদ্রিত হলে বলব যে, প্রচেষ্টাটি ব্যর্থ হ'ল। তাহলে, স্পষ্টতঃই পৃষ্ঠাপ্রতি অক্ষরসংখ্যা বিপুল পরিমাণ হবে এবং মুদ্রাকর ও মুদ্রণযন্ত্র ভালো হলে মুদ্রণভ্রান্তির সম্ভাবনা অর্থাৎ প্রচেষ্টায় সার্থকতালাভের সম্ভাবনা খুব অল্প হবে। তাছাড়া প্রচেষ্টাগুলিকে আমরা পরস্পর নির্ভরতাশূন্য ব'লেও মানতে পারি, কারণ মুদ্রণ কাজটি যেহেতু যন্ত্রসাহায্যে হচ্ছে এবং মুদ্রাকর প্রতিটি পৃষ্ঠাই যথাসাধ্য নির্ভুলভাবে ছাপবার চেষ্টা করছেন, কাজেই কোন একটি অক্ষর যদি হঠাৎ ভুল ছাপা হয়ে যায় তবে পার্শ্ববর্তী অন্যান্য অক্ষরগুলি সঙ্গে সঙ্গে ভুল ছাপা হবার সম্ভাবনার হ্রাসবৃদ্ধি হওয়ার কথা নয়। কাজেই ধরা যায় যে, এখানে অসংখ্য স্বনির্ভর বেরণুলীয় প্রচেষ্টা চলছে যাতে সার্থকতাসম্ভাবনা প্রতি প্রচেষ্টার জন্যে সমান এবং অতি ক্ষুদ্র। ফলে, আশা করা যায় যে, পৃষ্ঠাপ্রতি ভুল ছাপা

অক্ষরের পরিসংখ্যা-বিভাজন পোয়াসঁ বিভাজনের অনুরূপ হবে। এখানে পোয়াসঁ পূর্ণকাঙ্ক $m$-এর প্রাক্-কলক হিসেবে নেওয়া হবে পৃষ্ঠাপ্রতি গড় মুদ্রণ-প্রমাদজড়িত অক্ষরসংখ্যা, যার মান স্পষ্টতঃই খুব সামান্য হবে। কাজেই সবদিক বিবেচনা ক'রে এক্ষেত্রে পোয়াসঁ বিভাজনের সঙ্গে পরিদৃষ্ট পরিসংখ্যা-বিভাজনের সাযুজ্য খুবই প্রত্যাশিত।

আরও অনেকক্ষেত্রে পোয়াসঁ বিভাজনের সঙ্গে পরিলক্ষিত পরিসংখ্যা-বিভাজনের সাযুজ্য আশা করা যায়। যেমন, (1) দিনের কর্মচঞ্চল কয়েক ঘণ্টায় কোন টেলিফোন কেন্দ্রে আগত ডাকের পরিসংখ্যা-বিভাজন, (2) কোন ঘন তরল পদার্থে গুঁড়ো গুঁড়ো আটা জাতীয় পদার্থ ছড়িয়ে দিলে তার যে বিভাজন দৃষ্ট হয়, (3) ক্ষুদ্র ক্ষুদ্র সময়ের ফাঁকে ফাঁকে কোন রাস্তায় মোটর গাড়ী চলে যাওয়ার সংখ্যা, (4) কোন শহরের বিভিন্ন অঞ্চলে ছড়িয়ে পড়া বোমার খণ্ডের পরিসংখ্যা-বিভাজন ইত্যাদি।

## সারণী 8.6

### 345 জন ব্যক্তির দক্ষিণ হস্তের কব্জির শক্তির পরিসংখ্যা-বিভাজন

| দক্ষিণ-কব্জির শক্তি ( পাউণ্ডে ) | পরিসংখ্যা |
|---|---|
| 29·5— 39·5 | 1 |
| 39·5— 49·5 | 2 |
| 49·5— 59·5 | 12 |
| 59·5— 69·5 | 52 |
| 69·5— 79·5 | 99 |
| 79·5— 89·5 | 101 |
| 89·5— 99·5 | 55 |
| 99·5—109·5 | 17 |
| 109·5—119·5 | 5 |
| 119·5—129·5 | 1 |

এখন আমরা একটি প্রদত্ত পরিসংখ্যা-বিভাজনের সঙ্গে একটি নর্ম্যাল-বিভাজনের সাযুজ্য নিরূপণ দেখাবার চেষ্টা করব।

**উদা. 8.3** পূর্বপৃষ্ঠায় প্রদত্ত 8.6 সারণীটিতে কয়েক ব্যক্তির দক্ষিণ হাতের কব্জির শক্তি কতখানি তার একটি পরিসংখ্যা-বিভাজন দেওয়া হয়েছে। এর সঙ্গে একটি নর্ম্যাল বিভাজনের সাযুজ্য নিরূপণের চেষ্টা করা যাক।

এই সারণীতে প্রদর্শিত বিভাজনের পরিঘাতগুলি নির্ণয়ের জন্যে নীচে আর একটি সারণী গঠন করা হ'ল। নমুনাভিত্তিক পরিঘাতগুলিকে আমরা $\hat{\mu}'_r$, $\hat{\mu}_r$ ইত্যাদি চিহ্ন দিয়ে নির্দেশ করব।

## সারণী 8.7

### 8.6 সারণীতে পরিবেশিত তথ্যের ভিত্তিতে পরিঘাত নির্ণয়

| শ্রেণীমধ্যক $x$ | পরিসংখ্যা $f$ | $\xi=\frac{x-84.5}{10}$ | $f\xi$ | $f\xi^2$ | $f\xi^3$ | $f\xi^4$ | $f(\xi+1)^4$ |
|---|---|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| 34·5 | 1 | −5 | −5 | 25 | −125 | 625 | 256 |
| 44·5 | 2 | −4 | −8 | 32 | −128 | 512 | 162 |
| 54·5 | 12 | −3 | −36 | 108 | −324 | 972 | 192 |
| 64·5 | 52 | −2 | −104 | 208 | −416 | 832 | 52 |
| 74·5 | 99 | −1 | −99 | 99 | −99 | 99 | 0 |
| 84·5 | 101 | 0 | 0 | 0 | 0 | 0 | 101 |
| 94·5 | 55 | 1 | 55 | 55 | 55 | 55 | 880 |
| 104·5 | 17 | 2 | 34 | 68 | 136 | 272 | 1377 |
| 114·5 | 5 | 3 | 15 | 45 | 135 | 405 | 1280 |
| 124·5 | 1 | 4 | 4 | 16 | 64 | 256 | 625 |
| মোট | 345 | – | −144 | 656 | −702 | 4028 | 4925 |

## সারণী 8.8

| শ্রেণী | শ্রেণীমধ্যক $x$ | শ্রেণীসীমান্ত $y$ | $t = \frac{y-\bar{x}}{s}$ | $\Phi(t)$ | $\Delta\Phi(t)$ | $N\Delta\Phi(t)$ ( অখণ্ড সংখ্যায় ) | প্রদত্ত পরিসংখ্যা | $u = \frac{x-\bar{x}}{s}$ | $\phi(u)$ | $\frac{N}{s}\phi(u)$ |
|---|---|---|---|---|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| 29·5 – 39·5 | 34·5 | 29·5 | –3·867 | ·000055 | ·000893 | 0 | 1 | –3·487 | ·000913 | ·0236 |
| 39·5 – 49·5 | 44·5 | 39·5 | –3·106 | ·000948 | ·008566 | 3 | 2 | –2·726 | ·009712 | ·2547 |
| 49·5 – 59·5 | 54·5 | 49·5 | –2·345 | ·009514 | ·049397 | 17 | 12 | –1·965 | ·057872 | 1·5202 |
| 59·5 – 69·5 | 64·5 | 59·5 | –1·564 | ·058911 | ·152079 | 52 | 52 | –1·204 | ·193256 | 5·0753 |
| 69·5 – 79·5 | 74·5 | 69·5 | – ·803 | ·210990 | ·272260 | 94 | 99 | – ·443 | ·361654 | 9·4968 |
| 79·5 – 89·5 | 84·5 | 79·5 | – ·042 | ·483250 | ·280679 | 97 | 101 | ·318 | ·379270 | 9·9589 |
| 89·5 – 99·5 | 94·5 | 89·5 | ·719 | ·763929 | ·166634 | 57 | 55 | 1·079 | ·222894 | 5·8525 |
| 99·5 – 109·5 | 104·5 | 99·5 | 1·480 | ·930563 | ·056922 | 20 | 17 | 1·840 | ·073407 | 1·9272 |
| 109·5 – 119·5 | 114·5 | 109·5 | 2·241 | ·987487 | ·011172 | 4 | 5 | 2·601 | ·013548 | ·3545 |
| 119·5 – 129·5 | 124·5 | 119·5 | 3·002 | ·998659 | ·001251 | 0 | 1 | 3·362 | ·001401 | ·0368 |
| | | 129·5 | 3·763 | ·999910 | — | — | — | | | |

শার্লিয়ারের শুদ্ধি পরীক্ষা (Charlier's check) :

$$\sum f(\xi+1)^4 = \sum f\xi^4 + 4\sum f\xi^3 + 6\sum f\xi^2 + 4\sum f\xi + \sum f$$

প্রদত্ত রাশিমালার জন্যে স্পষ্টতঃই এ সম্পর্কটি সত্য।

আমরা লিখব $N = \sum f = 345$, $h = 10 =$ শ্রেণীঅন্তরের দৈর্ঘ্য। $\xi$ চলের পরিঘাতগুলিকে $\hat{\mu}'_r(\xi)$ ও $\hat{\mu}_r(\xi)$ দ্বারা চিহ্নিত ক'রে পাই

$$\hat{\mu}'_1(\xi) = -\text{·}4174,\ \hat{\mu}'_2(\xi) = 1\text{·}9014,\ \hat{\mu}'_3(\xi) = -2\text{·}0345,$$

$$\hat{\mu}'_4(\xi) = 11\text{·}6754.$$

তাহলে, $\bar{x} = 84\text{·}5 + 10 \times \mu'_1(\xi) = 80\text{·}326$,

এবং $\hat{\mu}_r = h^r\, \hat{\mu}_r(\xi)$—এই সূত্র সাহায্যে পাই

$$\hat{\mu}_2 = \hat{\sigma}^2 = s^2 = 100 \times 1\text{·}727,\ \hat{\mu}_3 = 1000 \times \text{·}201,\ \hat{\mu}_4 = 10000 \times 10\text{·}1749.$$

$$s = \hat{\sigma} = 13\text{·}142,\ \hat{\beta}_1 = \text{·}0078,\ \hat{\beta}_2 = 3\text{·}411$$

($\beta_1$-$\beta_1$) লেখচিত্রে (·0078, 3·411) বিন্দুটি স্থাপন ক'রে দেখা যায় যে, প্রদত্ত পরিসংখ্যা-বিভাজনের সঙ্গে পিয়ার্সনীয় সপ্তম প্রকার অথবা নর্ম্যাল-বিভাজনের ঘনিষ্ঠ সাযুজ্য থাকা সম্ভব। এখন নর্ম্যাল বিভাজনের সঙ্গে এই বিভাজনের সাযুজ্য কতখানি রয়েছে 8.8 সারণীতে দেখা যাক।

8.8 সারণীর উল্লম্বপঙ্‌ক্তি (7) ও (8) তুলনা ক'রে মোটামুটি ভালো মিল দেখা যায়। তাছাড়া পরিসংখ্যাঘনত্ব [ (8) ÷ 10 ] এবং উল্লম্ব পঙ্‌ক্তি (11)-এ প্রদর্শিত মানগুলির মধ্যেও পরস্পরিক মিল ভালোই দেখা যাচ্ছে। তাই বলা যায় যে, প্রদত্ত বিভাজনের সঙ্গে একটি নর্ম্যাল বিভাজনের মোটামুটি ভালো সাযুজ্য রয়েছে।

## 8.4 অনুশীলনী

8.1 ঔপপত্তিক ( তত্ত্বগত ) বিভাজন কাকে বলে? বিচ্ছিন্ন ও অবিচ্ছিন্ন উভয় প্রকার চলের সম্পর্কেই এ বিষয়ে আলোচনা কর। ঔপপত্তিক বিভাজনের সম্ভাবনা আদর্শ বলতে কী বোঝায়?

8.2 বাইনোমিয়াল বিভাজনের গড়কেন্দ্রিক পরিঘাত সম্পর্কিত পৌনঃপুনিকতা সূত্র ব্যবহার ক'রে ঐ বিভাজনের $\mu_2$, $\mu_3$ ও $\mu_4$-এর মান নির্ণয় কর।

8.3 বাইনোমিয়াল বিভাজন $b(x ; n, p)$-এর চিহ্ননিরপেক্ষ গড়বিচ্যুতি কত ?

[ উত্তর : $2npq\binom{n-1}{\mu-1}p^{\mu-1}\,q^{n-\mu}$,

যেখানে $\mu=[np]+1$ ]

8.4 দুটি উদাহরণ দিয়ে বুঝিয়ে দাও কোন্ ক্ষেত্রে কেন পোয়াসঁ বিভাজন কোন প্রদত্ত বিভাজনের সঙ্গে সাযুজ্যপূর্ণ হওয়া স্বাভাবিক।

8.5 পোয়াসঁ বিভাজন $P(x ; m)$-এর ভূয়িষ্ঠক নির্ণয় কর।

[ উত্তর : $[m]$ ; $m$ অখণ্ড হলে $m$ ও $m-1$ উভয়েই হবে ভূয়িষ্ঠক। অবশ্য এক্ষেত্রে বলা হবে যে ভূয়িষ্ঠকের অস্তিত্ব নেই। ]

8.6 একটি পুস্তিকায় মোট 30টি পৃষ্ঠা আছে। ভাল ক'রে নিরীক্ষণ ক'রে দেখা গেল যে তাতে 20টি পৃষ্ঠা সম্পূর্ণভাবে মুদ্রণ-প্রমাদ মুক্ত। এখন, যদি সমসম্ভব উপায়ে এর 5টি পৃষ্ঠা বেছে নেওয়া হয় তাহলে তাদের প্রত্যেকটিতেই কিছু না কিছু মুদ্রণ-প্রমাদ থাকার সম্ভাবনা কত ? [ উত্তর : $\binom{10}{5}\Big/\binom{30}{5}$ ]

8.7 দুই সহস্র সংখ্যক আলপিনের একটি বাক্সে মোট চল্লিশটি ত্রুটিযুক্ত আলপিন আছে। এই বাক্সটি থেকে গৃহীত 20টি আলপিনের মধ্যে অনধিক একটি ত্রুটিযুক্ত আলপিন পাবার সম্ভাবনা কত ? পোয়াসঁ বিভাজন অনুযায়ী এই সম্ভাবনার আসন্ন মান কত হবে ? বাইনোমিয়াল বিভাজন অনুযায়ী এই সম্ভাবনার আসন্নমানই বা কত ?

[ উত্তর : $$\sum_{x=0}^{1}\binom{40}{x}\binom{1960}{20-x}\Big/\binom{2000}{20} \,;\; \frac{7}{5}\,e^{-\frac{2}{5}}\,;$$

$$\sum_{x=0}^{1}\binom{20}{x}\left(\frac{1}{50}\right)^{x}\left(\frac{49}{50}\right)^{20-x}$$

8.8 $X$ যদি এমন একটি সম্ভাবনা চল হয় যার জন্যে

$$P[X=x]=f(x)=\tfrac{1}{2}g(x)+\tfrac{1}{2}h(x),\quad x=0, 1,\ldots, n,$$

$$g(x)=\binom{n}{x}p_1^{x}\,q_1^{n-x},\ 0<p_1, q_1<1,\ p_1+q_1=1$$

$$h(x)=\binom{n}{x}p_2^{x}\,q_2^{n-x},\ 0<p_2, q_2<1,\ p_2+q_2=1,$$

তাহলে $E(X)$ ও $V(X)$ এর মান $p_1, q_1, p_2, q_2$ ও $n$-এর মাধ্যমে প্রকাশ কর।

[ উত্তর : $\frac{1}{2}n(p_1+p_2)$ ; $\frac{n^2}{4}(p_1-p_2)^2+\frac{n}{2}(p_1+p_2)$

$$-\frac{n}{2}(p_1^2+p_2^2)].$$

8.9 দেখাও যে,

$$f(x)=\frac{a^p}{\Gamma(p)}e^{-ax}x^{p-1}, a>0, 0\leq x<\infty,$$

একটি সম্ভাবনা-ঘনত্ব-অপেক্ষক নির্দেশ করে। এর প্রথম চারটি পরিঘাত ও $\beta_1, \beta_2$-অঙ্ক নির্ণয় কর।

[ উত্তর : $E(X)=\mu'_1=\frac{p}{a}, \mu_2=\frac{p}{a^2}, \mu_3=\frac{2p}{a^3},$

$$\mu_4=\frac{3p^2+6p}{a^4}, \beta_1=\frac{4}{p}, \beta_2=3+\frac{6}{p};$$

ফলে যদি $p\to\infty$ হয়, তাহলে $\sqrt{\beta_1}\to 0$ এবং $\beta_2\to 3$.]

8.10 দেখাও যে,

$$f(x)=\frac{1}{B(m, n)}x^{m-1}(1-x)^{n-1}, 0<x<1 ; m, n>0$$
$$=0 \quad \text{অন্যথায়};$$

$$g(x)=\frac{1}{B(m, n)}\frac{x^{2n-1}}{(1+x^2)^{m+n}}, 0<x<\infty ; m, n>0$$
$$=0 \quad \text{অন্যথায়};$$

ও $$h(x)=\frac{1}{B(m, n)}\frac{x^{n-1}}{(1+x)^{m+n}}, 0<x<\infty ; m, n>0$$
$$=0 \quad \text{অন্যথায়};$$

—এই অপেক্ষকগুলি সম্ভাবনা-ঘনত্ব নির্দেশ ক'রে। এই বিভাজনগুলির প্রথম চারটি পরিঘাত ও $\beta_1, \beta_2$-অঙ্ক নির্ণয় কর।

[ উত্তর : $f(x)$-এর জন্যে $\mu'_1=\frac{m}{m+n}$

$$\mu_2=\frac{mn}{(m+n)^2(m+n+1)}, \text{ ইত্যাদি।}]$$

8.11 ধর, একটি সম্ভাবনা চল $X$-এর সম্ভাবনা-ভর-অপেক্ষক হচ্ছে $P[X=x]=f(x)=\theta^x(1-\theta)^{1-x}$, $x=0, 1$ ; $0<\theta<1$.

এই চলটির প্রথম চারটি পরিঘাত ও $\beta_1, \beta_2$-অঙ্ক নির্ণয় কর। এই

বিভাজনটির ভূয়িষ্ঠক কত হবে ? $\theta=\frac{1}{2}$ হলে দেখাও যে বিভাজনটির মধ্যমমান হচ্ছে 0.

[ উত্তর : এটি দ্বিপদ বিভাজনের একটি বিশেষ ক্ষেত্র ($n=1$). কাজেই পূর্বে আলোচিত বিষয়ের সাহায্যে নিজে নির্ণয় কর। $X$-এর ভূয়িষ্ঠক হবে 1 যদি $\theta<\frac{1}{2}$ হয় এবং এর ভূয়িষ্ঠক হবে 0 যদি $\theta<\frac{1}{2}$ হয়। ]

8.12 ধর, একটি সম্ভাবনা চল $X$-এর সম্ভাবনা-ভর-অপেক্ষক হচ্ছে

$$P[X=x]=f(x)=k\theta^x(1-\theta),\ 0<\theta<1\ ;\ x=0,1,2,\ldots$$

$E(X)$ ও $V(X)$-এর মান নির্ণয় কর।

$$\left[\text{উত্তর :}\quad E(X)=\frac{\theta}{1-\theta},\quad V(X)=\frac{\theta}{(1-\theta)^2}.\right]$$

8.13 $f(x)=q^xp,\ 0<p,q<1\ ;\ x=0,1,2,\ldots$

ও $g(x)=\binom{r+x-1}{x}p^rq^x,\ 0<p,q<1\ ;\ x=0,1,2,\ldots$

সম্ভাবনা-ভর-অপেক্ষকযুক্ত সম্ভাবনা-বিভাজনের গড় ও ভেদমান নির্ণয় কর।

$$\left[\text{উত্তর :}\quad \frac{q}{p},\frac{q}{p^2}\ ;\ \frac{rq}{p},\frac{rq}{p^2}\right]$$

8.14 যদি একটি অবিচ্ছিন্ন সম্ভাবনা চল $X$-এর সম্ভাবনা-ঘনত্ব-অপেক্ষক হয়

$$f(x)=\frac{1}{\theta}e^{-\frac{x}{\theta}},\ \theta>0\ ;\ 0\leq x<\infty,$$

তাহলে $X$-এর গড় ও ভেদমান এবং গড়কেন্দ্রিক চিহ্ননিরপেক্ষ গড়বিচ্যুতি নির্ণয় কর।

$$\left[\text{উত্তর :}\quad \theta,\ \theta^2,\ \frac{2\theta}{e}\right]$$

8.15 একটি পরীক্ষায় পাশ করার জন্যে শতকরা অন্ততঃ 30 নম্বর, দ্বিতীয় বিভাগে পাশ করার জন্যে শতকরা অন্ততঃ 45 নম্বর ও প্রথম বিভাগে পাশ করার জন্যে শতকরা অন্ততঃ 60 নম্বর পাওয়া দরকার হয়। যদি ঐ পরীক্ষায় প্রাপ্ত নম্বরের বিভাজনকে নর্ম্যাল প্রকৃতির ব'লে ধরা যায় এবং ঐ পরীক্ষায় শতকরা 22 জন অকৃতকার্য হয় ও শতকরা 15 জন প্রথম বিভাগে উত্তীর্ণ হয়, তাহলে তাতে দ্বিতীয় বিভাগে উত্তীর্ণ পরীক্ষার্থীর শতকরা হার কত ?

[ উত্তর : 30% ]

8.16 একটি শহরের অধিবাসীদের দৈহিক ওজন সম্পর্কে তথ্য নিয়ে জানা গেছে যে ওজনের চতুর্থক বিচ্যুতি হচ্ছে 4 কিলোগ্রাম এবং তাদের শতকরা 20 ভাগের ওজন 18 কিলোগ্রামের কম। ওজনের বিভাজন নর্ম্যাল ধরে নিয়ে নির্ণয় কর তাদের মধ্যে শতকরা কতজনের ওজন 30 কিলোগ্রামের চেয়ে বেশী হবে। [ উত্তর : 12% ( আসন্নভাবে ) ]

8.17 দেখাও যে $f(x)=\frac{e^{-m}m^x}{(1-e^{-m})x!}$; $x=1, 2, 3, \ldots ; m>0$,

একটি সম্ভাবনা ভর অপেক্ষক নির্দেশ করে। এই বিভাজনের গড় কত ?

$\left[\text{উত্তর : } \frac{m}{1-e^{-m}}\right]$

8.18. $f(x)=y_0\left(1-\frac{x^2}{a^2}\right)^m, \; -a<x<a$

দ্বারা নির্দেশিত সম্ভাবনা বিভাজনের ভেদমান কত ? $\left[\text{উত্তর : } \frac{a^2}{2m+3}\right]$

8.19 দেখাও যে, $h(x)=\frac{1}{\sigma}\sqrt{\frac{2}{\pi}}\, e^{-\frac{x^2}{2\sigma^2}}, \; 0<x<\infty$

একটি সম্ভাবনা ঘনত্ব অপেক্ষক। বিভাজনটির গড় ও ভেদমান কত ?

$\left[\text{উত্তর : } \sigma\sqrt{\frac{2}{\pi}}, \; \sigma^2\left(1-\frac{2}{\pi}\right)\right]$

8.20 রাশিবিজ্ঞানে নর্ম্যাল বিভাজনের উপযোগিতা আলোচনা কর।

## 8.5 নির্দেশিকা

1. Elderton W. P. and Johnson, N. L. *Systems of Frequency Curves.* Cambridge University Press, 1969.

2. Feller. W. *An Introduction to Probability Theory and Its Applications, Vol. 1.* Asia Publishing House, 1960.

3. Goon, A. M, Gupta, M. K, and Dasgupta, B. *Fundamentals of Statistics, Vol. 1.* The World Press, Ltd. Calcutta, 1975.

4. Uspensky, J. V. *Introduction to Mathematical Probability.* Mc. Graw Hill, 1937.

5. Weatherburn, C. E. *A First Course in Mathematical Statistics.* Cambridge University Press, 1947.

6. Yule, G. U and Kendall, M. G. *An Introduction to The Theory of Statistics.* Charles Griffin and Co. Ltd. London, 1955.

# 9 গুণলক্ষণের সংস্রব
## ( Association of Attributes )

**9.1 গুণলক্ষণের যৌথ-বিভাজন:** প্রদত্ত কিছুসংখ্যক ব্যষ্টির একটিমাত্র লক্ষণ ( গুণগত অথবা পরিমাণগত ) সম্পর্কে সংগৃহীত রাশিতথ্য নিয়েই এ পর্যন্ত আলোচনা করা হয়েছে। অনেক সময় একই সঙ্গে একাধিক লক্ষণ সংক্রান্ত তথ্য সংগ্রহ করার প্রয়োজন হতে পারে, বিশেষ ক'রে যখন সংশ্লিষ্ট লক্ষণগুলির মধ্যে পারস্পরিক কোনরূপ সম্পর্ক আছে কি না তা নির্ধারণ করার প্রশ্ন ওঠে। এই উদ্দেশ্যে একাধিক গুণলক্ষণ সম্পর্কে সংগৃহীত রাশিতথ্য বিশ্লেষণের বিষয়টির বিস্তারিত আলোচনা করা হবে বর্তমান পরিচ্ছেদে—পরবর্তী পরিচ্ছেদ নেওয়া হবে একাধিক চলের প্রসঙ্গটি।

যুগপৎ একাধিক গুণলক্ষণ সম্পর্কে প্রদত্ত কিছুসংখ্যক ব্যষ্টির পরিসংখ্যা-বিভাজন আগের মতই মিলচিহ্নের সাহায্যে পাওয়া যায়। নীচের উদাহরণটি লক্ষ্য কর। এখানে 'স্কুল ফাইন্যাল পরীক্ষার ফল' এই গুণলক্ষণটির তিনটি রূপ নেওয়া হয়েছে সারি বরাবর এবং 'প্রি ইউনিভার্সিটি ( পি. ইউ. ) পরীক্ষার ফল' এই গুণলক্ষণটির তিনটি রূপ নেওয়া হয়েছে স্তম্ভ বরাবর। এইভাবে ছকটি $3 \times 3 = 9$টি কোষে বিভক্ত হয়েছে। যে ছাত্রটি স্কুল ফাইন্যাল প্রথম বিভাগে এবং পি. ইউ. দ্বিতীয় বিভাগে পাশ করেছে তার জন্য প্রথম সারির দ্বিতীয় কোষে একটি মিলচিহ্ন নেওয়া হয়েছে। এইভাবে 1493 জন ছাত্রছাত্রীকে ( সকলেই এই দুটি পরীক্ষায় পাশ করেছে ধ'রে নিয়ে ) তাদের এই দুটি পরীক্ষার ফলাফলের ভিত্তিতে 9টি কোষের কোন না কোনটিতে মিলচিহ্নের সাহায্যে সন্নিবেশিত ক'রে মিলচিহ্নগুলি গণনা ক'রে তাদের পরিসংখ্যা বিভাজন পাওয়া গেছে 9.1 সারণীতে।

সাধারণভাবে মনে কর $A$ এবং $B$ এই দুটি গুণলক্ষণের যথাক্রমে $A_1, A_2, \ldots, A_r$ এবং $B_1, B_2, \ldots, B_s$—এই $r$ এবং $s$টি বিভিন্ন রূপ আছে। গুণলক্ষণ দুটির সম্পর্কে দ্বিধারা পরিসংখ্যা বিভাজনে $f_{ij}$ দ্বারা নির্দেশ করা যাক যুগপৎ $A$-এর $i$-তম রূপ এবং $B$-এর $j$-তম রূপ-সম্বলিত ব্যষ্টির সংখ্যা ($i = 1, 2, \ldots, r$ ; $j = 1, 2, \ldots, s$)। $f_{ij}$-কে বলা হয় $(i, j)$-তম কোষের পরিসংখ্যা

(cell-frequency)। স্পষ্টতঃই এখানে মোট কোষের সংখ্যা $r \times s$. মনে কর $A$-এর $i$-তম রূপের জন্য বিভিন্ন কোষপরিসংখ্যাগুলির যোগফল $f_{io}$ দ্বারা এবং $B$-এর

### সারণী 9.1

### স্কুল ফাইন্যাল ও প্রি. ইউনিভার্সিটী পরীক্ষার ফলাফলের ভিত্তিতে 1493 জন ছাত্রছাত্রীর পরিসংখ্যা বিভাজন

| স্কুল ফাইনাল | প্রি. ইউনিভার্সিটি | | | |
|---|---|---|---|---|
| | প্রথম বিভাগ | দ্বিতীয় বিভাগ | তৃতীয় বিভাগ | মোট |
| প্রথম বিভাগ | 175 | 54 | 3 | 232 |
| দ্বিতীয় বিভাগ | 44 | 319 | 79 | 442 |
| তৃতীয় বিভাগ | 9 | 91 | 719 | 819 |
| মোট | 228 | 464 | 801 | 1,493 |

$j$-তম রূপের জন্য বিভিন্ন কোষপরিসংখ্যাগুলির যোগফল $f_{oj}$ দ্বারা চিহ্নিত করা হল। অর্থাৎ,

$$f_{io} = \sum_{j=1} f_{ij} \text{ এবং } f_{oj} = \sum_{i=1} f_{ij} \qquad (9.1)$$

এখন
$$n = \sum_{i=1} \sum_{j=1} f_{ij} = \sum_{i=1} f_{io} = \sum_{j=1} f_{oj} = f_{oo} \quad \cdots (9.2)$$

হলে, $rs$টি কোষপরিসংখ্যা এবং সর্বমোট পরিসংখ্যা $f_{oo}$—$A$ ও $B$ এই দুটি গুণলক্ষণের যৌথ ( পরিসংখ্যা ) বিভাজন [joint (frequency) distribution] সূচিত করে। $f_{oo}$ সহযোগে $A$-এর $r$ টি প্রান্তিক পরিসংখ্যা [marginal frequency] $f_{io}$ $(i=1, 2..., r)$ এবং $B$-এর $s$ টি প্রান্তিক পরিসংখ্যা $f_{oj}$ $(j=1, 2,..., s)$ যথাক্রমে সূচিত করে $A$ ও $B$-র প্রান্তিক ( পরিসংখ্যা ) বিভাজন [marginal (frequency) distribution]। $A$-এর একটি নির্দিষ্ট রূপ $A_i$-এর জন্য $s$-টি কোষপরিসংখ্যা $f_{ij}$ $(j=1, 2,...s)$ সংশ্লিষ্ট প্রান্তিক পরিসংখ্যা $f_{io}$ সহযোগে যে ( পরিসংখ্যা ) বিভাজনটি সূচিত করে তাকে বলা হয় $A$-এর $i$-তম

রূপের জন্য $B$-এর সর্তাধীন ( পরিসংখ্যা ) বিভাজন [conditional (frequency) distribution]। $B$-এর বিভিন্ন নির্দিষ্ট রূপের জন্য $A$-রও $s$টি বিভিন্ন সর্তাধীন বিভাজন পাওয়া যায়।

ওপরের সারণীতে 228, 464, 801 এবং 1493 ও 232, 442, 819 এবং 1493 যথাক্রমে পি. ইউ.-এর ফলাফল ও স্কু. ফা.-এর ফলাফলের প্রান্তিক ( পরিসংখ্যা ) বিভাজন নির্দেশ করে। 175, 54, 3 এবং 232 সূচিত করে স্কু. ফা.-এর প্রথম বিভাগে উত্তীর্ণ ছাত্রদের পি. ইউ.-এর ফলাফলের সর্তাধীন বিভাজন। তেমনি পি. ইউ.-এ তৃতীয় বিভাগে উত্তীর্ণ ছাত্রছাত্রীদের স্কু. ফা.-এর ফলাফলের সর্তাধীন বিভাজন পাওয়া যায় 3, 79, 719 এবং 801 থেকে।

দুই-এর বেশী গুণলক্ষণের সম্পর্কেও তথ্যসংগ্রহ করা হয় অনেক সময়। 9.2 সারণীটি তিনটি গুণলক্ষণের যৌথ পরিসংখ্যা বিভাজনের উদাহরণ।

সাধারণভাবে $A$, $B$ এবং $C$ এই তিনটি গুণলক্ষণের বিভিন্ন রূপ $A_i$, $B_j$ এবং $C_k$ $(i=1, 2, \ldots, r\,;\ j=1, 2, \ldots, s\,;\ k=1, 2, \ldots, t)$ দ্বারা এবং কোষপরিসংখ্যাগুলি ( সংখ্যায় $r \times s \times t$ টি ) $f_{ijk}$ দ্বারা নির্দেশ করা হয়। এখানে মনে করা যাক,

$$\left.\begin{aligned} \sum_{i=1}^{r} f_{ijk} &= f_{ojk} \\ \sum_{j=1}^{s} f_{ijk} &= f_{iok} \\ \sum_{k=1}^{t} f_{ijk} &= f_{ijo} \end{aligned}\right\} \qquad \ldots (9.3)$$

$$\left.\begin{aligned} \sum_i \sum_j f_{ijk} &= \sum_i f_{iok} = \sum_j f_{ojk} = f_{ook} \\ \sum_i \sum_k f_{ijk} &= \sum_i f_{ijo} = \sum_k f_{ojk} = f_{ojo} \\ \sum_j \sum_k f_{ijk} &= \sum_j f_{ijo} = \sum_k f_{iok} = f_{ijo} \end{aligned}\right\} \qquad \ldots (9.4)$$

এবং $n \cdots \sum_{i}\sum_{j}\sum_{k} f_{ijk} = \sum_{i}\sum_{j} f_{ijo} = \sum_{j}\sum_{k} f_{ojk}$

$$= \sum_{i}\sum_{k} f_{iok} = \sum_{i} f_{ioo} = \sum_{j} f_{ojo} = \sum_{k} f_{ook} = f_{ooo}. \quad \cdots (9.5)$$

$f_{ooo}$ সহযোগে $f_{ioo}$ $(i=1, 2, \cdots, r)$, $f_{ojo}$ $(j=1, 2, \cdots, s)$ এবং $f_{ook}$ $(k=1, 2, \cdots, t)$ যথাক্রমে $A$, $B$ ও $C$-এর প্রান্তিক বিভাজন সূচিত করে। আর এক ধরণের প্রান্তিক বিভাজন পাওয়া যায় $f_{ijo}$, $f_{iok}$ এবং $f_{ojk}$ থেকে —এগুলো হ'ল যথাক্রমে $A$ ও $B$-এর, $A$ ও $C$-এর এবং $B$ ও $C$-এর যৌথ

**সারণী 9.2** [ কাল্পনিক ]

সাক্ষরতা, প্রগতিশীলতা এবং উচ্চাকাঙ্ক্ষা অনুসারে
374 জন ব্যক্তির পরিসংখ্যা বিভাজন

| | স্বাক্ষর | | | নিরক্ষর | | | মোট |
|---|---|---|---|---|---|---|---|
| | উচ্চাকাঙ্ক্ষী | উচ্চাকাঙ্ক্ষী নয় | মোট | উচ্চাকাঙ্ক্ষী | উচ্চাকাঙ্ক্ষী নয় | মোট | |
| প্রগতিশীল | 130 | 33 | 163 | 64 | 31 | 95 | 258 |
| গোঁড়া | 29 | 43 | 72 | 10 | 34 | 44 | 116 |
| মোট | 159 | 76 | 235 | 74 | 65 | 139 | 374 |

প্রান্তিক বিভাজন। সর্তাধীন বিভাজনও দু'ধরনের হবে এক্ষেত্রে। প্রথমতঃ, $A$ ও $B$-এর প্রদত্ত রূপের জন্য $C$-এর, $B$ ও $C$-এর প্রদত্ত রূপের জন্য $A$-এর এবং $A$ ও $C$-এর প্রদত্ত রূপের জন্য $B$-এর সর্তাধীন বিভাজন হতে পারে। এবং দ্বিতীয়তঃ, $A$-র প্রদত্ত রূপের জন্য $B$ ও $C$-এর, $B$-এর প্রদত্ত রূপের জন্য $A$ ও $C$-এর এবং $C$-এর প্রদত্ত রূপের জন্য $A$ ও $B$-এর যৌথ সর্তাধীন বিভাজন পাওয়া যায়।

### 9.2 গুণলক্ষণের যৌথ বিভাজন সংক্রান্ত রাশি-তথ্যের সামঞ্জস্য (consistency of data on joint distribution of attributes) :

আগেই বলা হয়েছে, যথাক্রমে $r$ ও $s$টি রূপবিশিষ্ট দুটি গুণলক্ষণের যৌথ বিভাজনে $rs$টি কোষ-পরিসংখ্যা, $r+s$টি প্রান্তিক পরিসংখ্যা এবং একটি সর্বমোট

পরিসংখ্যা পাওয়া যায়। তবে এদের সবগুলি পরস্পর নিরপেক্ষ নয়, কারণ এগুলি (9.1) ও (9.2) সূত্রে প্রদত্ত সমীকরণগুলির দ্বারা পরস্পর সম্বন্ধযুক্ত। বাস্তবিকপক্ষে এই $rs+r+s+1$টি বিভিন্ন ধরনের পরিসংখ্যার মধ্যে পরস্পর নিরপেক্ষ পরিসংখ্যার সর্বাধিক সংখ্যা হ'ল $rs$. স্পষ্টতঃই, যে কোন $rs$টি বিভিন্ন ধরনের পরিসংখ্যা পরস্পর নিরপেক্ষ নাও হতে পারে। যেমন $2 \times 2$ পরিসংখ্যা সারণীতে কোষপরিসংখ্যাগুলি $f_{11}$, $f_{12}$, $f_{21}$ এবং $f_{22}$ দ্বারা, প্রান্তিক পরিসংখ্যাগুলি $f_{10}, f_{20}, f_{01}$ ও $f_{02}$ দ্বারা এবং সর্বমোট পরিসংখ্যাটি $f_{00}$ দ্বারা নির্দেশ করা হলে, $f_{11}$, $f_{12}$, $f_{21}$ ও $f_{22}$ ; $f_{11}$, $f_{10}$ $f_{21}$ ও $f_{20}$ ; বা $f_{11}$, $f_{01}$, $f_{20}$ ও $n$—এই গুচ্ছগুলির প্রত্যেকটিতে প্রদত্ত পরিসংখ্যাগুলি পরস্পর নিরপেক্ষ। কিন্তু $f_{11}, f_{12}, f_{10}$ এবং $f_{01}$ ; বা $f_{01}, f_{10}, f_{02}$ এবং $f_{20}$—এই গুচ্ছদুটির পরিসংখ্যাগুলি পরস্পর নিরপেক্ষ নয়, কেননা, $f_{11}+f_{12}=f_{10}$ এবং $f_{01}+f_{02}=f_{10}+f_{20}$. স্পষ্টতঃই, একটি $r \times s$ সারণীর যে কোন $rs$টি পরস্পর নিরপেক্ষ পরিসংখ্যা ( কোষ-, প্রান্তিক- অথবা সর্বমোট ) দেওয়া থাকলে অন্যান্য পরিসংখ্যাগুলি সহজেই নির্ণয় করা যায় এবং যৌথ-বিভাজনটি সম্পূর্ণভাবে নির্দিষ্ট হয়। উদাহরণস্বরূপ একটি $2 \times 2$ সারণীতে $f_{11}=4$, $f_{01}=5$, $f_{20}=4$ এবং $f_{00}=14$ দেওয়া থাকলে অন্যান্যগুলি হবে $f_{21}=f_{01}-f_{11}=1$, $f_{02}=f_{00}-f_{01}=9$, $f_{22}=f_{20}-f_{21}=3$, $f_{10}=f_{00}-f_{20}=10$, $f_{12}=f_{10}-f_{11}=6$.

অনুরূপভাবে যথাক্রমে $r$, $s$ ও $t$টি রূপবিশিষ্ট তিনটি গুণলক্ষণের যৌথ বিভাজনে বেশীপক্ষে $rst$ টি পরিসংখ্যা পরস্পর নিরপেক্ষ হতে পারে এবং যে কোন $rst$-টি পরস্পর নিরপেক্ষ পরিসংখ্যা দেওয়া থাকলে যৌথ বিভাজনটি সম্পূর্ণভাবে চিহ্নিত করা সম্ভব।

অনেক সময় একাধিক গুণলক্ষণের যৌথ বিভাজন সম্পর্কে মাত্র আংশিক রাশিতথ্য দেওয়া থাকে। প্রদত্ত আংশিক বা সম্পূর্ণ রাশিতথ্যগুলিকে সমঞ্জস (consistent) বলা হবে যদি তথ্যগুলি পরস্পরবিরোধী না হয়। যেমন $f_{11}=17$, $f_{10}=13$—দুইটি গুণলক্ষণের যুগ্মবিভাজন সম্পর্কে এই তথ্য দুটি পরস্পরবিরোধী, তাই অসমঞ্জস, কেননা $f_{11} > f_{10}$ হওয়া সম্ভব নয়।

স্পষ্টতঃই একাধিক গুণলক্ষণের যৌথ বিভাজন সম্বন্ধে প্রদত্ত একপ্রস্থ রাশিতথ্য সমঞ্জস হওয়ার জন্য আবশ্যিক এবং পর্যাপ্ত (necessary and sufficient) সর্ত হ'ল, লব্ধ কোষ পরিসংখ্যা এবং প্রান্তিক পরিসংখ্যাগুলির প্রত্যেকটিকে অঋণাত্মক ( অর্থাৎ, ধনাত্মক অথবা শূন্য ) হতে হবে। এই সর্তের বিচারেই

প্রদত্ত রাশিতথ্যের অন্তর্নিহিত অসামঞ্জস্য নির্ণয় করা চলে। নীচের উদাহরণটি দেখ।

**উদা. 9.1** মোট 1,000টি কৃষিজমির মধ্যে সারপ্রদত্ত, সেচযুক্ত, ধানী, সেচযুক্ত সারপ্রদত্ত, সারপ্রদত্ত ধানী এবং সেচযুক্ত ধানী জমির সংখ্যা যথাক্রমে 510, 490, 427, 189, 140 এবং 85. দেখাও যে তথ্যগুলির মধ্যে অসামঞ্জস্য আছে।

এখানে সংশ্লিষ্ট গুণলক্ষণ হ'ল তিনটি—যথাক্রমে সেচব্যবস্থা, সারপ্রয়োগ এবং উৎপন্ন কৃষিদ্রব্যের বিচারে জমির প্রকৃতি—এবং প্রতিটি গুণলক্ষণের দুটি ক'রে রূপ আছে। $A$ এবং $\alpha$ দ্বারা যথাক্রমে সেচযুক্ত ও সেচবিহীন, $B$ ও $\beta$ দ্বারা যথাক্রমে সারপ্রদত্ত ও সারবিহীন এবং $C$ ও $\gamma$ দ্বারা যথাক্রমে ধানী ও অন্যান্য জমি নির্দেশ করা হলে

প্রদত্ত তথ্য থেকে আমরা পাই,

$$f_A = 490, f_B = 510, f_C = 427$$

$$f_{AB} = 189, f_{BC} = 140, f_{AC} = 85, n = 1{,}000.$$

এখন, $f_{\alpha\beta\gamma} = n - f_{ABC} - f_{AB\gamma} - f_{\alpha BC} - f_{A\alpha C}$

$$- f_{A\beta\gamma} - f_{\alpha B\gamma} - f_{\alpha\beta C}$$

$$= n - f_{ABC} - (f_{AB\gamma} + f_{AB\gamma})$$

$$- (f_{\alpha BC} + f_{\alpha B\gamma}) - (f_{A\beta C} + f_{\alpha\beta C})$$

$$= n - f_{ABC} - f_{A\gamma} - f_{\alpha B} - f_{\beta C}$$

$$= n - f_{ABC} - (f_A - f_{AC}) - (f_B - f_{AB}) - (f_C - f_{BC})$$

$$= n - f_{ABC} - f_A - f_B - f_C + f_{AB} + f_{BC} + f_{AC}$$

$$= 1{,}000 - f_{ABC} - 490 - 510 - 427 + 189 + 140 + 85$$

$$= -f_{ABC} - 13.$$

স্পষ্টতঃই, যেহেতু $f_{ABC}$ অ-ঋণাত্মক, $f_{\alpha\beta\gamma}$ অবশ্যই ঋণাত্মক হবে, যা সম্ভব নয়। সুতরাং প্রদত্ত তথ্যগুলির মধ্যে অসামঞ্জস্য রয়েছে।

## 9.3 সংস্রব এবং অনপেক্ষতা (association and independence) :

### 9.3.1 2 × 2 সারণীর ক্ষেত্রে :

আগেই বলা হয়েছে, একাধিক গুণলক্ষণ সংক্রান্ত তথ্যসংগ্রহের একটা উদ্দেশ্য হ'ল সংশ্লিষ্ট লক্ষণগুলির মধ্যে কোন পারস্পরিক সম্বন্ধ আছে কি না তা নির্ণয়

করা। আলোচনার সুবিধার জন্য ধরা যাক, $A$ এবং $B$ এই গুণলক্ষণদুটির প্রত্যেকটির দুটি ক'রে রূপ আছে—$A$ ও $\alpha$ এবং $B$ ও $\beta$. মনে কর $A$ ও $B$ লক্ষণদুটির উপস্থিতি এবং $\alpha$ ও $\beta$ লক্ষণদুটির অনুপস্থিতি সূচিত করে। সুতরাং $f_{AB}$, $f_{A\beta}$, $f_{\alpha\beta}$ ও $f_{\alpha\beta}$—এগুলি হ'ল কোষপরিসংখ্যা, $f_A$ ও $f_\alpha$ $A$-লক্ষণটির এবং $f_B$ ও $f_\beta$ $B$-লক্ষণটির প্রান্তিক পরিসংখ্যা এবং ধরা যাক সর্বমোট পরিসংখ্যা হ'ল $n$. স্পষ্টতঃই

$$\left.\begin{array}{ll} f_{AB}+f_{A\theta}=f_A & f_{AB}+f_{\alpha B}=f_B \\ f_{\alpha B}+f_{\alpha\beta}=f_\alpha & f_{A\beta}+f_{\alpha\beta}=f_\alpha \end{array}\right\} \quad \cdots (9.6)$$

এবং
$$\left.\begin{array}{l} f_{AB}+f_{A\beta}+f_{\alpha B}+f_{\alpha\beta} \\ =f_A+f_\alpha \\ =f_B+f_\beta=n. \end{array}\right\} \quad \cdots (9.7)$$

মনে কর, প্রান্তিক পরিসংখ্যাগুলির মান শূন্যেতর। তাহলে $f_{AB}/f_B$ এবং $f_{A\beta}/f_\beta$-এই অনুপাতদুটি সূচিত করে যেসব ব্যষ্টির মধ্যে $B$-লক্ষণটি যথাক্রমে উপস্থিত এবং অনুপস্থিত তাদের মধ্যে $A$ লক্ষণাক্রান্ত ব্যষ্টিগুলির অনুপাত।

সুতরাং
$$\frac{f_{AB}}{f_B}=\frac{f_{A\beta}}{f_B} \quad \cdots (9.8)$$

হওয়ার অর্থ $B$-এর উপস্থিতি অথবা অনুপস্থিতি $A$-এর উপস্থিতির আনুপাতিক পরিসংখ্যাকে আদৌ প্রভাবিত করে না। এক্ষেত্রে বলা হয়, $A$ ও $B$ এই গুণলক্ষণদুটি রাশিবিজ্ঞানগতভাবে অনপেক্ষ বা অনধীন (statistically independent)। অন্যথায় বলা হয়ে থাকে $A$ ও $B$-এর মধ্যে সংস্রব (association) রয়েছে, বা $A$ ও $B$ পরস্পর সংস্রবযুক্ত (associated)। সুতরাং $A$ ও $B$ পরস্পর অনপেক্ষ হওয়ার সর্ত হ'ল

$$\frac{f_{AB}}{f_B}=\frac{f_{AB}}{f_\beta}$$

বা
$$f_{AB}/f_B=(f_{AB}+f_{AB})/(f_B+f_\beta)=f_A/n$$

অর্থাৎ,
$$f_{AB}=f_A.f_B/n \quad \cdots (9.9)$$

স্পষ্টতঃই, (9.8) থেকে পাওয়া যায়

$$\left.\begin{aligned} f_{A\beta} &= \frac{f_A . f_\beta}{n} \\ f_{\alpha B} &= \frac{f_\alpha . f_B}{n} \\ \text{এবং} \quad f_{\alpha\beta} &= \frac{f_\alpha . f_\beta}{n} \end{aligned}\right\} \quad \cdots \quad (9.9a)$$

(9.9) সমীকরণটিকে $A$ ও $B$-এর অনপেক্ষতা-নির্দেশী মূলসূত্র বলা যেতে পারে, কেননা, (9.9$a$) সূত্রে প্রদত্ত সমীকরণগুলি (9.9) সূত্র থেকেই উদ্ভূত। লক্ষণীয়, এক বা একাধিক প্রান্তিক পরিসংখ্যার মান শূন্য হলেও সূত্রটি ব্যবহার করায় কোন অসুবিধা নেই।

(9.9) সমীকরণটি সত্য না হলে, নিম্নলিখিত অসমতা দুটির একটি সত্য হবে

$$f_{AB} > \frac{f_A . f_B}{n} \quad \cdots \quad (9.10a)$$

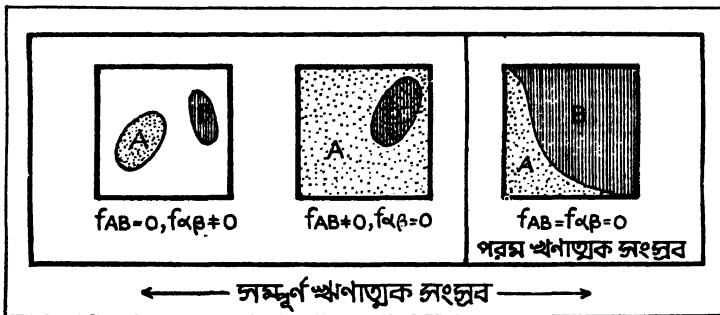
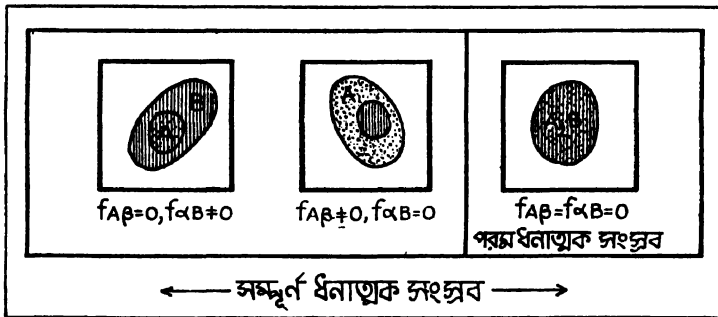
$$f_{AB} < \frac{f_A . f_B}{n} \quad \cdots \quad (9.10b)$$

প্রথমোক্ত ক্ষেত্রে $A$ এবং $B$ অনপেক্ষ হলে লক্ষণদুটির যতসংখ্যক ব্যষ্টির মধ্যে একত্রে উপস্থিত থাকার কথা, তার থেকে বেশী সংখ্যক ব্যষ্টিতে প্রকৃতপক্ষে উপস্থিত রয়েছে। আর দ্বিতীয় ক্ষেত্রে ঘটেছে ঠিক উল্টোটি। উভয় ক্ষেত্রেই $A$ ও $B$-এর মধ্যে সংশ্রব বর্তমান, তবে সংশ্রবের প্রকৃতি দুটি ক্ষেত্রে ভিন্ন। (9.10$a$) এবং (9.10$b$) সূত্রদুটি যথাক্রমে ধনাত্মক (positive) এবং ঋণাত্মক (negative) সংশ্রবের সূচক।

আবার $f_{A\beta}=0$ এবং $f_{\alpha B}=0$ হলে, অর্থাৎ $A$ লক্ষণাক্রান্ত সমস্ত ব্যষ্টিই $B$ লক্ষণাক্রান্ত, ও $B$ লক্ষণাক্রান্ত সমস্ত ব্যষ্টিই $A$ লক্ষণাক্রান্ত হলে বলা হয় $A$ এবং $B$-এর মধ্যে রয়েছে পরম ধনাত্মক সংশ্রব (absolute positive association)। অনুরূপভাবে পরম ঋণাত্মক সংশ্রবের (absolute negative association) সর্ত হ'ল $f_{AB}=0$ এবং $f_{\alpha\beta}=0$—অর্থাৎ $A$ লক্ষণাক্রান্ত কোন ব্যষ্টিই $B$ লক্ষণাক্রান্ত হবে না এবং $A$ লক্ষণাক্রান্ত নয় এমন কোন ব্যষ্টিতেই $B$ লক্ষণটি অনুপস্থিত থাকবে না।

$f_{A\beta}=0$ এবং $f_{\alpha B}=0$ এই দুটি সর্তের অন্ততঃ একটি পালিত হলে পাওয়া যায় সম্পূর্ণ ধনাত্মক সংশ্রব (complete positive association) এবং $f_{AB}=0$

ও $f_{\alpha\beta}=0$—এই সর্তদুটির অন্ততঃ একটি পালিত হলে পাওয়া যায় সম্পূর্ণ ঋণাত্মক সংশ্রব (complete negative association)। পরম সংশ্রব এবং সম্পূর্ণ সংশ্রব হ'ল নিখুঁত সংশ্রবের (perfect association) দুটি বিভিন্ন রূপ। স্পষ্টতঃই দুটি গুণলক্ষণের সংশ্রব পরম হলে তা সম্পূর্ণও বটে, কিন্তু উল্টোটি সত্য নাও হতে পারে।

চিত্র 9.1 থেকে বিভিন্ন ধরনের নিখুঁত সংশ্রব সম্বন্ধে ধারণা পাওয়া যাবে।

চিত্র 9.1
বিভিন্ন ধরনের নিখুঁত সংশ্রব

## 9.3.2 r×s সারণীর ক্ষেত্রে :

এক্ষেত্রে $A$ ও $B$ এই দুইটি লক্ষণের বিভিন্ন রূপগুলি যথাক্রমে $A_1, A_2, \ldots, A_r$ এবং $B_1, B_2, \ldots, B_s$ দ্বারা চিহ্নিত ক'রে 9.1 অনুচ্ছেদের মতো $f_{ij}$, $f_{io}$ এবং $f_{oj}$ $(i=1, 2, \ldots, r\,;\ j=1, 2, \ldots, s)$ দ্বারা যথাক্রমে কোষ-পরিসংখ্যা, এবং $A$ ও $B$-এর প্রান্তিক পরিসংখ্যা সূচিত করা যেতে পারে।

স্পষ্টতঃই, $f_{io} > 0, f_{oj} > 0$ এই স্বীকরণসাপেক্ষে $j$-এর প্রতিটি মানের জন্য

$$\frac{f_{1j}}{f_{10}}=\frac{f_{2j}}{f_{20}}=\dots=\frac{f_{rj}}{f_{ro}},$$ হলে

বা $i$ ও $j$-এর প্রতিটি মানের জন্য

$$\frac{f_{ij}}{f_{io}}=\frac{\sum_{i=1}^{r} f_{ij}}{\sum_{i=1}^{r} f_{io}}=\frac{f_{oj}}{n}$$ হলে

অর্থাৎ, $f_{ij}=f_{io}\times f_{oj}/n$ ··· (9.11)

হলে, $A$ এবং $B$-কে রাশিবিজ্ঞানসম্মতভাবে অনপেক্ষ বলা হবে। অন্যথায় অন্ততঃ একটি $(i, j)$-এর জন্যও সমীকরণটি সত্য না হলে বলা হবে $A$ এবং $B$-এর মধ্যে সংস্রব রয়েছে। $A$ এবং $B$ পরস্পর অনপেক্ষ কি না বিচার করার জন্য (9.11) সমীকরণে বর্ণিত সর্তগুলি যাচাই করা হয়। আপাতদৃষ্টিতে সর্তগুলি সংখ্যায় মোট $rs$ টি হলেও এদের মধ্যে মাত্র $(r-1)(s-1)$টি পরস্পর নিরপেক্ষ, কারণ $f_{ij}$, $f_{io}$ এবং $f_{oj}$-এই পরিসংখ্যাগুলি (9.1) এবং (9.2) সমীকরণে প্রদত্ত কয়েকটি সর্তের (constraint) অধীন।

$r\times s$ সারণীর ক্ষেত্রে ধনাত্মক এবং ঋণাত্মক সংস্রবের মধ্যে পার্থক্য নির্ধারণ সাধারণতঃ খুব একটা অর্থবহ হয় না। অবশ্য সংশ্লিষ্ট লক্ষণ দুটির বিভিন্ন রূপগুলি কোনভাবে মানের ক্রমানুসারে সাজানো সম্ভব হলে (যেমন 9.1 সারণীতে) ধনাত্মক সংস্রব এবং ঋণাত্মক সংস্রবের ব্যাখ্যানও সম্ভব। 9.1 সারণীতে পরিবেশিত রাশিতথ্য থেকে এক নজরেই বলা যায় স্কুল ফাইন্যাল এবং প্রি. ইউনিভার্সিটি পরীক্ষার ফলাফলের মধ্যে ধনাত্মক সংস্রব রয়েছে।

এই পরিস্থিতিতে $r=s$ হলে এবং $A$ ও $B$-এর বিভিন্ন রূপগুলি একইভাবে মানের ক্রমানুসারে সাজানো সম্ভব হলে নিখুঁত ধনাত্মক সংস্রবের সর্ত হবে এই যে, প্রতিটি $i$ ও $j$-এর জন্য

$$fij=0,\ i\neq j \qquad \cdots \quad (9.12a)$$

এবং নিখুঁত ঋণাত্মক সংস্রবের সর্ত হবে এই যে, প্রতিটি $i$ ও $j$-এর জন্য

$$fij=0,\ j\neq r-i+1. \qquad \cdots \quad (9.12b)$$

## 9.4 সংস্রব-মাপক (measures of association) :

### 9.4.1 আদর্শ সংস্রব-মাপকের ধর্মাবলী :

দুটি গুণলক্ষণ $A$ এবং $B$ পরস্পর অনপেক্ষ নয়, কেবলমাত্র এই জ্ঞানটুকুই অনেক সময় পর্যাপ্ত হয় না, লক্ষণ দুটির মধ্যে কী ধরনের (ধনাত্মক অথবা ঋণাত্মক) এবং কতখানি সংস্রব বর্তমান তাও জানা দরকার হতে পারে। এইসব প্রশ্নের উত্তর পেতে হলে উপযুক্ত সংস্রব মাপকের কথা ভাবতে হবে। এখন প্রশ্ন হ'ল : এই ধরনের একটি আদর্শ সংস্রব-মাপকের কোন্ কোন্ ধর্ম থাকবে ? প্রথম কথা, মাপকটির প্রকৃতি সহজে বোধগম্য হতে হবে। দ্বিতীয়তঃ, স্পষ্টতঃই মাপকটির মান অনপেক্ষতা, ধনাত্মক সংস্রব এবং ঋণাত্মক সংস্রবের ক্ষেত্রে যথাক্রমে শূন্য, ধনাত্মক এবং ঋণাত্মক হওয়া প্রয়োজন। তৃতীয়তঃ, মোট পরিসংখ্যা $n$-এর ওপর মাপকটির নির্ভরশীল হওয়া উচিত নয় আদৌ। চতুর্থতঃ, ব্যাখ্যানের সুবিধার জন্য মাপকটির মান সাধারণতঃ একটি নির্ধারিত মানসীমার, যেমন $\pm 1$-এর, মধ্যে সীমাবদ্ধ থাকতে হবে। নিখুঁত ঋণাত্মক সংস্রবের ক্ষেত্রে সম্ভাব্য লঘিষ্ঠ মান থেকে শুরু ক'রে সংস্রবের মাত্রা বাড়ার সঙ্গে সঙ্গে মাপকটিরও মান ক্রমশঃ বাড়া উচিত। এইভাবে বাড়তে বাড়তে এটি অনপেক্ষতার ক্ষেত্রে শূন্য এবং নিখুঁত ধনাত্মক সংস্রবের ক্ষেত্রে নির্ধারিত সীমার গরিষ্ঠ মানসম্পন্ন হওয়া উচিত।

### 9.4.2 2×2 সারণীর ক্ষেত্রে :

9.3.1 অনুচ্ছেদের আলোচনা থেকে স্বভাবতঃই

$$\delta_{AB} = f_{AB} - \frac{f_A f_B}{n} \quad \cdots \quad (9.13)$$

—প্রকাশনটিকে এক্ষেত্রে সংস্রব-মাপকের ভিত্তি হিসাবে নেওয়ার কথা মনে হবে। $\delta_{AB}$-ভিত্তিক নিম্নলিখিত মাপকটিকে **সংস্রবাঙ্ক** (coefficient of association) বলা হয় :

$$Q_{AB} = \frac{n\delta_{AB}}{f_{AB} . f_{\alpha\beta} + f_{A\beta} . f_{\alpha B}}.$$

এখন $n\delta_{AB} = nf_{AB} - f_A f_B$.

$$= f_{AB}(f_{AB} + f_{\alpha B} + f_{AB} + f_{\alpha\beta}) - (f_{AB} + f_{AB})(f_{A\beta} + f_{\alpha B})$$

$$= f_{AB} . f_{\alpha\beta} - f_{A\beta} . f_{\alpha B}$$

সুতরাং $$Q_{AB} = \frac{f_{AB} . f_{\alpha\beta} - f_{A\beta} . f_{\alpha B}}{f_{AB} . f_{\alpha\beta} + f_{A\beta} . f_{\alpha B}}. \quad \cdots \quad (9.14)$$

$A$ ও $B$ পরস্পর অনপেক্ষ হলে $\delta_{AB}=0$, সুতরাং $Q_{AB}=0$. $A$ ও $B$-এর মধ্যে সম্পূর্ণ ঋণাত্মক সংস্রব থাকলে $f_{AB}$ এবং $f_{\alpha\beta}$-এ দুটির মধ্যে অন্ততঃ একটির মান শূন্য, সুতরাং $f_{AB}.f_{\alpha\beta}=0$ অর্থাৎ, $Q_{AB}=-1$. পক্ষান্তরে সম্পূর্ণ ধনাত্মক সংস্রবের ক্ষেত্রে $f_{A\beta}$ ও $f_{\alpha B}$-এর মধ্যে অন্ততঃ একটির মান শূন্য, সুতরাং $f_{A\beta}.f_{\alpha B}=0$ অর্থাৎ $Q_{AB}=+1$. আবার কোষ-পরিসংখ্যাগুলির মান অঋণাত্মক, সুতরাং $Q_{AB}$-এর গরিষ্ঠ এবং লঘিষ্ঠ মান যথাক্রমে $+1$ এবং $-1$.

দ্বিতীয় আর একটি সংস্রব-মাপক হ'ল **সংশ্লেষনাঙ্ক** (coefficient of colligation)। এটির সূত্র :

$$Y_{AB}=\frac{\sqrt{f_{AB}.f_{\alpha\beta}}-\sqrt{f_{A\beta}f_{\alpha B}}}{\sqrt{f_{AB}.f_{\alpha\beta}}+\sqrt{f_{A\beta}f_{\alpha B}}} \quad \cdots \quad (9.15)$$

$$=\frac{\sqrt{ad}-\sqrt{bc}}{\sqrt{ad}+\sqrt{bc}}, \quad \cdots \quad (9.16)$$

$f_{AB}=a$, $f_{\alpha B}=b$, $f_{A\beta}=c$, $f_{\alpha\beta}=d$ ধরে নিয়ে। স্পষ্টতঃই $Q_{AB}$ এবং $Y_{AB}$-এর সাধারণ ধর্মগুলি একই।

সহজেই দেখানো যায়, $\frac{2Y_{AB}}{1+Y^2_{AB}}=Q_{AB}$. $\quad \cdots \quad (9.17)$

এ দুটি ছাড়াও আরও একটি সংস্রব-মাপকের প্রচলন রয়েছে। এটি হ'ল

$$V_{AB}=\frac{n\delta_{AB}}{\sqrt{f_Af_Bf_\alpha f_\beta}}=\frac{f_{AB}f_{\alpha\beta}-f_{A\beta}f_{\alpha B}}{\sqrt{f_Af_Bf_\alpha f_\beta}} \quad \cdots \quad (9.18)$$

স্পষ্টতঃই, $A$ ও $B$ পরস্পর অনপেক্ষ হলে $V_{AB}=0$. এটিরও সম্ভাব্য মানসীমা $\pm 1$. এখন দেখা যাক, কোন্ কোন্ পরিস্থিতিতে মাপকটির মান $+1$ কিংবা $-1$ হয়।

(9.16) সূত্রে প্রদত্ত প্রতীকচিহ্ন ব্যবহার ক'রে

$$V_{AB}=\frac{ad-bc}{\{(a+b)(c+d)(a+c)(b+d)\}^{\frac{1}{2}}} \quad \cdots \quad (9.19)$$

সুতরাং $V_{AB}=\pm 1$ হওয়ার আবশ্যিক এবং পর্যাপ্ত সর্ত হ'ল

$$(ad-bc)^2=(a+b)(c+d)(a+c)(b+d)$$

বা, $$a^2(bc+bd+cd)+b^2(ac+ad+cd)+c^2(ab+ad+bd)$$
$$+d^2(ab+ac+bc)+4\ abcd=0$$

এখন লক্ষ্য কর কেবলমাত্র যদি $a$, $b$, $c$ এবং $d$-এর যে কোন দুটির মান শূন্য হয় তবেই বামপক্ষটির মান শূন্য হবে।

$a=b=0$, $c=d=0$, $a=c=0$ এবং $b=d=0$—এগুলির কোনটিই সত্য হওয়া সম্ভব নয়, কেননা প্রান্তিক পরিসংখ্যাগুলির মান শূন্যেতর ধ'রে নেওয়া হয়েছে। সুতরাং বামপক্ষটি শূন্য হবে একমাত্র যদি $b=c=0$ কিংবা $a=d=0$ হয়। কিন্তু $b=c=0$ হওয়ার অর্থ $A$ এবং $B$-এর মধ্যে পরম ধনাত্মক সংশ্রব রয়েছে এবং সেক্ষেত্রে $V_{AB}$-এর মানও ধনাত্মক ( 9.19 দ্রষ্টব্য ), সুতরাং $+1$. পক্ষান্তরে $a=d=0$ হলে $A$ ও $B$-এর পরম ঋণাত্মক সংশ্রব সূচিত হয় এবং তখন $V_{AB}$-এর মান দাঁড়ায় $-1$.

সুতরাং দেখা যাচ্ছে $V_{AB}$-এর সম্ভাব্য মানসীমা $\pm1$ এবং কেবলমাত্র পরম সংশ্রবের ক্ষেত্রেই এটি প্রান্তিক মানদুটি গ্রহণ করে।

**উদা. 9.2** সাম্প্রতিক একটি সমীক্ষায় ভারতের কোন একটি অঞ্চলের 2483 জন অধিবাসী সম্বন্ধে নিম্নলিখিত তথ্য পাওয়া গেছে।

**সারণী 9.3**

| মদ্যপান করে কি না | মানসিক স্বাস্থ্য | | |
|---|---|---|---|
| | ভাল | ভাল নয় | মোট |
| ক'রে | 761 | 255 | 1016 |
| করে না | 993 | 474 | 1467 |
| মোট | 1754 | 729 | 2483 |

মানসিক স্বাস্থ্য এবং মদ্যপান যথাক্রমে $A$ ও $B$ দ্বারা চিহ্নিত করা যাক এখানে,

$$Q_{AB}=\frac{761\times474-255\times993}{761\times474+255\times993}=\frac{107499}{613929}=\cdot1751$$

$$Y_{AB}=\frac{\sqrt{761\times474}-\sqrt{255\times993}}{\sqrt{761\times474}+\sqrt{255\times993}}=\frac{97\cdot3889}{1103\cdot7971}$$
$$=\cdot0882$$

সুতরাং দেখা যাচ্ছে, মদ্যপান এবং মানসিক স্বাস্থ্যের মধ্যে ধনাত্মক সংশ্রব রয়েছে, যদিও সংশ্রবের মাত্রা খুবই কম।

### 9.4.3. $r\times s$ সারণীর ক্ষেত্রে

এক্ষেত্রেও $\delta_{ij}=\dfrac{f_{io}\times f_{oj}}{n}$ ... (9.20)

—এই প্রকাশনটি প্রয়োজনীয় সংশ্রব মাপকের ভিত্তি হিসাবে নেওয়া যেতে পারে। আসলে $r\times s$ সারণীর ক্ষেত্রে

$$\chi^2{}_{AB}=\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{\delta_{ij}{}^2}{(f_{io}\times f_{oj})/n}$$

$$=n\sum\sum_{j}\frac{f_{ij}{}^2}{f_{io}\times f_{oj}}-n \qquad \cdots \quad (9.21)$$

—এই রাশিটিকে মাপক হিসাবে নেওয়া হয়। স্পষ্টতঃই $\chi^2{}_{AB}=0$ হবে কেবলমাত্র যদি প্রতিটি $(i, j)$-এর জন্য $\delta_{ij}=0$ হয়, অর্থাৎ $A$ ও $B$ পরস্পর অনপেক্ষ হয়। $A$ এবং $B$ এর মধ্যে সংশ্রবের মাত্রা যত বেশী হবে ( ধনাত্মক বা ঋণাত্মক যে দিকেই হোক ) $\chi^2{}_{AB}$-এর মানও তত বেশী হবে। তবে সংশ্রব মাপক হিসাবে $\chi^2{}_{AB}$-এর প্রধান ত্রুটি, এটি $n$-এর ওপর একান্ত নির্ভরশীল এবং অন্ততঃ তাত্ত্বিক বিচারে এর মান সীমাহীনভাবে বৃহৎ হতে পারে। সুতরাং $\chi^2{}_{AB}$-এর লব্ধ মান থেকে সংশ্রবের মাত্রা সম্বন্ধে বিশেষ কিছু জানা যায় না।

পিয়ার্সনের (Pearson)-এর **সম্বন্ধাঙ্ক** (coefficient of contingency)

$$C_{AB}=\sqrt{\frac{\chi^2{}_{AB}}{n+\chi^2{}_{AB}}} \qquad \cdots \quad (9.22)$$

কিন্তু এই ত্রুটি থেকে মুক্ত। তবে এটির আর একটি ত্রুটি হ'ল, এর সম্ভাব্য সর্বোচ্চ মান 1-এর থেকে কম, কারণ যেহেতু $\chi^2{}_{AB}\geqslant 0, n>0$, সুতরাং $\chi^2{}_{AB}<n+\chi^2{}_{AB}$, অর্থাৎ নিখুঁত সংশ্রবের ক্ষেত্রেও এর মান 1-এর সমান হয় না। একটি উদাহরণ নেওয়া যাক। মনে কর $A$ এবং $B$-এর প্রত্যেকটির $r$-টি ক'রে বিভিন্ন রূপ আছে এবং $A$ ও $B$-র মধ্যে রয়েছে নিখুঁত ধনাত্মক সংশ্রব অর্থাৎ (9.12$a$) অনুসারে

$f_{ij}\neq 0$ যদি $i=j$ হয়, এবং $f_{ij}=0$ যদি $i\neq j$ হয়। এক্ষেত্রে,

$$\chi^2{}_{AB}=n\sum_{i}\sum_{j}\frac{f_{ij}{}^2}{f_{io}\cdot f_{oj}}-n$$

$$= n \sum_i \frac{f_{ij}^2}{f_{io} \cdot f_{oj}} - n, \text{ কারণ } f_{io} = f_{oi} = f_{ij}$$

$$= n(r-1),$$

অর্থাৎ, $C_{AB} = \sqrt{\dfrac{n(r-1)}{n(r-1)+n}} = \sqrt{\dfrac{r-1}{r}} < 1.$

চুপ্রো (Tchuprow) প্রদত্ত আর একটি সংস্রব মাপক, যথা

$$T_{AB} = \left\{ \frac{\chi^2_{AB}}{n\sqrt{(r-1)(s-1)}} \right\}^{\frac{1}{2}} \qquad \cdots \quad (9.23)$$

অবশ্য $C_{AB}$-এর এই ত্রুটি থেকে মুক্ত। $r \times r$ সারণীতে নিখুঁত সংস্রবের ক্ষেত্রে এটির মান স্পষ্টতঃই 1. অবশ্য $r \neq s$ হলে $T_{AB}$-এর ঊর্ধ্বসীমা সম্পর্কে বিশেষ কিছু জানা যায় না।

$r \times s$ সারণীতে আলোচ্য সবকটি মাপকই ধনাত্মক এবং ঋণাত্মক সংস্রবের মধ্যে ( যেসব ক্ষেত্রে অর্থবহ ) পার্থক্য নির্দেশ করতে পারে না—মাপকগুলির মান সবসময়ই ধনাত্মক।

**উদা. 9.3.** 9.1 সারণীতে প্রদত্ত রাশিতথ্যের ক্ষেত্রে বিভিন্ন সংস্রব মাপকগুলির মান নির্ণয় করা যাক।

| $(i, j)$ | $f_{ij}$ | $f_{io} \times f_{oj}$ | $(2)^2/(3)$ |
|---|---|---|---|
| (1) | (2) | (3) | (4) |
| (1, 1) | 175 | 52896 | 0·57897 |
| (1, 2) | 54 | 100796 | 0·01921 |
| (1, 3) | 3 | 186732 | 0·00043 |
| (2, 1) | 44 | 107648 | 0·02709 |
| (2, 2) | 319 | 205088 | 0·49618 |
| (2, 3) | 79 | 380016 | 0·02179 |
| (3, 1) | 9 | 185832 | 0·00005 |
| (3, 3) | 91 | 354042 | 0·01763 |
| (3, 3) | 719 | 656019 | 0·78803 |
| মোট | 1,493 | — | 1·94938 |

$$\chi^2_{AB} = 1493(1·94938 - 1)$$
$$= 1417·42434.$$
$$C_{AB} = \sqrt{1417·42434/2910·42434}$$
$$= 0·69787.$$
$$T_{AB} = \sqrt{1417·42434/1417·2434 \times 2}$$
$$= 0·68898.$$

সুতরাং লক্ষণদুটির মধ্যে সংশ্রবের পরিমাণ খুব বেশী।

## 9.5 যুগ্ম, বহুল এবং আংশিক সংশ্রব (joint, multiple and partial association) :

অনেক সময় প্রদত্ত কিছু সংখ্যক ব্যষ্টির জন্য তিন বা ততোধিক গুণলক্ষণের ওপর তথ্য আহরণের প্রয়োজন হতে পারে আগেই বলা হয়েছে। 9.1 অনুচ্ছেদে তিনটি গুণলক্ষণের যৌথ বিভাজনের বিষয় আলোচিত হয়েছে। আমাদের বর্তমান আলোচনা তিনটি গুণলক্ষণের মধ্যেই সীমিত রাখা হবে। সংখ্যাটি তিন অতিক্রম করলেও প্রদত্ত বিভিন্ন সংজ্ঞা এবং সূত্রগুলি মোটামুটিভাবে প্রযোজ্য থাকবে।

ধরা যাক, $A, B$ এবং $C$ এই তিনটি গুণলক্ষণের যথাক্রমে $r\ (A_1, A_2, \ldots, A_r)$ $s\ (B_1, B_2, \ldots, B_s)$ এবং $t\ (C_1, C_2, \ldots, C_t)$ টি বিভিন্ন রূপ আছে। 9.1 অনুচ্ছেদের সঙ্কেতচিহ্নগুলি ব্যবহার ক'রে $f_{ijk}$ দ্বারা কোষ-পরিসংখ্যাগুলি এবং $f_{ioo}, f_{ojo}, f_{ook}, f_{ijo}, f_{iok}$ এবং $f_{ojk}$ দ্বারা প্রান্তিক পরিসংখ্যাগুলি চিহ্নিত করা হ'ল। এইসব পরিসংখ্যা (9.3)–(9.5) সমীকরণে প্রদত্ত বিভিন্ন সর্তের অধীন। একই ধরনের যুক্তি প্রয়োগ ক'রে দেখানো যায়, যদি প্রতিটি $i, j$ এবং $k$-এর জন্য

$$f_{ijk} = \frac{f_{ioo} \times f_{ojo} \times f_{ook}}{n^2} \qquad \cdots \quad (9.24)$$

হয়, তাহলে $A$, $B$ এবং $C$ পরস্পর অনপেক্ষ হবে। (9.24)-এর মোট $rst$ টি সমীকরণের মধ্যে মাত্র $(r-1)(s-1)+(s-1)(t-1)+(t-1)(r-1)$ $+(r-1)(s-1)(t-1) = rst - r - s - t + 2$টি বীজগাণিতিক বিচারে পরস্পর অনধীন। যে কোন একটি $(i, j, k)$-এর জন্য (9.24) সত্য না হলে বুঝতে হবে

$A$, $B$ ও $C$ **যৌথভাবে সংস্রবযুক্ত** (jointly associated)। যৌথ সংস্রবের মাত্রা নিরূপণের জন্য $C_{AB}$-এর প্রতিরূপ

$$C_{ABC}=\sqrt{\frac{\chi^2_{ABC}}{n+\chi^2_{ABC}}} \qquad (9.25)$$

ব্যবহার করা হয়। এক্ষেত্রে

$$\chi^2_{ABC}=\sum_i\sum_j\sum_k\left(f_{ijk}-\frac{f_{ioo}.f_{ojo}.f_{ook}}{n^2}\right)^2\Bigg/\frac{f_{ioo}.f_{ojo}.f_{ook}}{n^2}. \qquad (9.26)$$

সাধারণতঃ তিনটি গুণলক্ষণের যৌথবিভাজনের ক্ষেত্রে আমরা আগ্রহী হই এদের মধ্যে বিশেষ একটি, ধরা যাক $A$, একত্রে অন্যদুটির সম্পর্কে অনপেক্ষ কি না তা জানায়। এই উদ্দেশ্যে প্রথমে $r\times s\times t$ সারণীটিকে একটি $r\times st$ সারণীতে রূপান্তরিত করা হয়—একদিকে $A$-এর $r$টি রূপ এবং অন্যদিকে $B$ ও $C$-এর বিভিন্ন রূপগুলির সম্ভাব্য $st$ টি জুটি নিয়ে। এখন প্রতিটি $(i, j, k)$-এর জন্য

$$f_{ijk}=\frac{f_{ioo}\times f_{ojk}}{n} \qquad \cdots \quad (9.27)$$

সর্তটি পালিত হলে $A$-কে একত্রে $B$ ও $C$-এর সম্পর্কে অনপেক্ষ বলা হবে। অন্যথায় $A$ একত্রে $B$ ও $C$-এর সঙ্গে সংস্রবযুক্ত এবং এই সংস্রবকে বলা হয় $B$ ও $C$-এর সঙ্গে $A$-র **বহুল সংস্রব** (multiple association)। স্পষ্টতঃ, (9.27) সূত্রে বীজগাণিতিক বিচারে পরস্পর অনধীন সমীকরণের সংখ্যা $(r-1)(st-1)$.

বহুল সংস্রব পরিমাপের জন্য (9.22) এবং (9.23)-এর অনুরূপ দুটি মাপক ব্যবহার করা যেতে পারে। এগুলি হ'ল

$$C_{A.BC}=\sqrt{\frac{\chi^2_{A.BC}}{n+\chi^2_{A.BC}}} \qquad (9.28)$$

এবং $$T_{A.BC}=\sqrt{\frac{\chi^2_{A.BC}}{n\sqrt{(r-1)(st-1)}}} \qquad (9.29)$$

এখানে, $$\chi^2_{A.BC}=\sum_i\sum_j\sum_k\left(f_{ijk}-\frac{f_{ioo}\times f_{ojk}}{n}\right)^2\Bigg/f.$$

$$=n\sum_i\sum_j\sum_k\frac{f_{ijk}^2}{f_{ioo}\times f_{ojk}}-n. \qquad \cdots \quad (9.30)$$

অনুরূপভাবে $C_{B.CA}$ বা $T_{B.CA}$ এবং $C_{C.AB}$ বা $T_{C.AB}$-এর সাহায্যে যথাক্রমে $C$ ও $A$-র সঙ্গে $B$-র এবং $A$ ও $B$-র সঙ্গে $C$-র বহুল সংস্রব মাপা যেতে পারে।

অনেক সময় আবার এমন হতে পারে যে $A$ ও $B$ প্রত্যেকে তৃতীয় আর একটি লক্ষণ $C$-র সঙ্গে সংশ্রবযুক্ত। সেক্ষেত্রে $A$ ও $B$-র মধ্যে সংশ্রবের প্রকৃতি ও মাত্রা, উভয় লক্ষণের সঙ্গে $C$-এর সংশ্রবের প্রকৃতি ও মাত্রা দ্বারা প্রভাবান্বিত হওয়া খুবই সম্ভব। সুতরাং এই ধরনের পরিস্থিতিতে $A$ ও $B$-এর মধ্যে প্রকৃত সংশ্রবের মাত্রা এবং প্রকৃতির ওপর আলোকপাত করার জন্য $C$-র প্রতিটি রূপের জন্য আলাদা আলাদা ভাবে $A$ ও $B$-র মধ্যে সংশ্রব নির্ণয় করা হয়। $C$-র নির্দিষ্ট $k$-তম রূপের জন্য এই সংশ্রবকে বলা হয় $C_k$ এর উপস্থিতিতে $A$ ও $B$-র মধ্যে **আংশিক সংশ্রব** (partial association)। $2 \times 2$ বা $r \times s$ সারণীর ক্ষেত্রে $A$ এবং $B$-র মধ্যে সংশ্রবের মাত্রা নির্ণয়ের সময় আমরা $A$ ও $B$-র সঙ্গে সংশ্রবযুক্ত সম্ভাব্য অন্যান্য লক্ষণগুলি উপেক্ষা করেছিলাম, তাই এইসব ক্ষেত্রে $A$ ও $B$ র সংশ্রবকে **সামগ্রিক সংশ্রব** (total association) বলা যেতে পারে।

$2 \times 2 \times 2$ সারণীর জন্য আংশিক সংশ্রবাঙ্কের একটি সরলীকৃত রূপ পাওয়া যায়। উদাহরণ 9.1 এর সংকেতচিহ্নগুলি ব্যবহার ক'রে আংশিক সংশ্রবাঙ্কের সূত্রগুলি নিম্নলিখিতভাবে লেখা যেতে পারে :

$$Q_{AB.C} = \frac{f_C \cdot \delta_{AB.C}}{f_{ABC} \cdot f_{\alpha\beta C} + f_{A\beta C} \cdot f_{\alpha BC}}$$

$$= \frac{f_{ABC} \cdot f_{\alpha\beta C} - f_{A\beta C}\, f_{\alpha BC}}{f_{ABC} \cdot f_{\alpha\beta C} - f_{A\beta C} \cdot f_{\alpha BC}} \quad \cdots \quad (9.31a)$$

এবং $$Q_{AB.\gamma} = \frac{f_\gamma \cdot \delta_{AB\gamma}}{f_{AB\gamma}\, f_{\alpha\beta\gamma} + f_{A\beta\gamma} \cdot f_{\alpha B\gamma}}$$

$$= \frac{f_{AB\gamma} \cdot f_{\alpha\beta\gamma} - f_{A\beta\gamma} \cdot f_{\alpha B\gamma}}{f_{AB\gamma} \cdot f_{\alpha\beta\gamma} + f_{A\beta\gamma} \cdot f_{\alpha B\gamma}} \quad \cdots \quad (9.31b)$$

এখানে, $$\delta_{AB.C} = f_{ABC} - \frac{f_{AC} f_{BC}}{f_C} \quad \cdots \quad (9.32a)$$

এবং $$\delta_{AB.\gamma} = f_{AB\gamma} - \frac{f_{A\gamma} f_{B\gamma}}{f_\gamma}. \quad \cdots \quad (9.32b)$$

$Q_{AB C}$ এবং $Q_{AB.\gamma}$-এর সাধারণ ধর্মগুলি $Q_{AB}$-এর সাধারণ ধর্মগুলির অনুরূপ।

এখন, $\delta_{AB.C}+\delta_{AB.\gamma}$

$$=f_{AB}-\frac{f_{AC}f_{BC}}{f_C}-\frac{f_{A\gamma}.f_{B\gamma}}{f_\gamma}$$

$$=\left(f_{AB}-\frac{f_A.f_B}{n}\right)-\left(\frac{f_{AC}.f_{BC}}{f_C}+\frac{f_{A\gamma}.f_{B\gamma}}{f_\gamma}-\frac{f_A.f_B}{n}\right)$$

$$=\delta_{AB}-\frac{nf_\gamma.f_{AC}.f_{BC}+nf_C.f_{A\gamma}.f_{B\gamma}-f_A.f_B.f_C.f_\gamma}{nf_C.f_\gamma}$$

$$=\delta_{AB}-\frac{n}{f_C.f_\gamma}\left(f_{AC}-\frac{f_A.f_C}{n}\right)\left(f_{BC}-\frac{f_B.f_C}{n}\right)$$

$$=\delta_{AB}-\frac{n}{f_C.f_\gamma}\,\delta_{AC}.\delta_{BC}$$

অর্থাৎ, $$\delta_{AB}=\delta_{AB.C}+\delta_{AB.\gamma}+\frac{n}{f_Cf_\gamma}\,\delta_{AC}.\delta_{BC} \qquad (9.33)$$

(9.33) থেকে দেখা যাচ্ছে, $\delta_{AB.C}$ এবং $\delta_{AB.\gamma}$-এর মান শূন্য হলেও $\delta_{AB}$-র মান শূন্যেতর হতে পারে—অর্থাৎ, অন্য একটি লক্ষণ $C$-র দুটি বিভিন্ন রূপের জন্য আলাদা আলাদা ভাবে $A$ এবং $B$ পরস্পর অনপেক্ষ হলেও লক্ষণদুটির মধ্যে সামগ্রিক সংশ্রব থাকা সম্ভব। আবার $\delta_{AB.C}$ এবং $\delta_{AB.\gamma}$-এর মান শূন্য না হয়েও $\delta_{AB}$-র মান শূন্য হতে পারে, অর্থাৎ সামগ্রিক সংশ্রবাঙ্কের বিচারে আপাতদৃষ্টিতে দুটি লক্ষণকে পরস্পর অনপেক্ষ মনে হলেও প্রকৃতপক্ষে তা কৃত্রিম হওয়া সম্ভব—এদের ওপর তৃতীয় কোন লক্ষণের প্রভাব হয়তো এই আপাত অনপেক্ষতার জন্য দায়ী।

সুতরাং সাধারণভাবে সংশ্রবাঙ্কের মান থেকে দুটি গুণলক্ষণের পারস্পরিক অনপেক্ষতা সম্বন্ধে খুব সাবধানে সিদ্ধান্ত নেওয়া প্রয়োজন। লক্ষণ দুটির মধ্যে প্রকৃতই কোন সংশ্রব আছে কি না বিচার করতে হলে এদের সঙ্গে সংশ্রবযুক্ত সম্ভাব্য অন্য এক বা একাধিক লক্ষণের বিভিন্ন রূপগুলির প্রতিটি সম্ভাব্য জুটির জন্য আলাদা আলাদা ভাবে লক্ষণদুটির মধ্যে আংশিক সংশ্রবের পরিমাণ নির্ধারণ করতে হবে। এই সব অন্যান্য লক্ষণের বিভিন্ন রূপের নির্দিষ্ট বিভিন্ন দুটির জন্য যদি লক্ষণদুটি সংশ্রবযুক্ত দেখা যায়, তবেই এদের মধ্যে যথার্থ সংশ্রব আছে বলা যাবে।

**উদা. 9.4** 9.2 সারণীতে প্রদত্ত রাশিতথ্যের জন্য বিভিন্ন ধরনের সংশ্রব মাপকের মান নির্ণয় করা যাক।

এখানে $A$ = সাক্ষরতা, $B$ = প্রগতিশীলতা এবং $C$ = উচ্চাকাঙ্ক্ষা ধ'রে নিয়ে পাওয়া যায়,

$$f_A = 235, \quad f_B = 258, \quad f_C = 233, \quad f_\alpha = 139, \quad f_\beta = 116,$$
$$f_\gamma = 141, \; f_{BC} = 194, \; f_{\beta C} = 39, \quad f_{B\gamma} = 64 \text{ এবং } f_{\beta\gamma} = 77.$$

$A$, $B$ ও $C$-র মধ্যে যৌথ সংস্রব মাপনার জন্য প্রথমতঃ 9.4 সারণীতে প্রদর্শিত ছকে অঙ্কপাতন করা যাক।

**সারণী 9.4**

| কোষ $(i, j, k)$ | কোষ-পরিসংখ্যা $f_{ijk}$ | $f_{ioo} \times f_{ojo} \times f_{ook}$ | $(2)^2/(3)$ |
|---|---|---|---|
| (1) | (2) | (3) | (4) |
| $ABC$ | 130 | 14126790 | ·0011963 |
| $AB\gamma$ | 33 | 8548830 | ·0001274 |
| $A\beta C$ | 29 | 6351580 | ·0001324 |
| $A\beta\gamma$ | 43 | 3843660 | ·0004811 |
| $\alpha BC$ | 64 | 8355846 | ·0004902 |
| $\alpha B\gamma$ | 31 | 5056542 | ·0001901 |
| $\alpha\beta C$ | 10 | 3756892 | ·0000266 |
| $\alpha\beta\gamma$ | 34 | 2273484 | ·0005085 |
| মোট | 374 | — | ·0031526 |

সুতরাং $\chi^2_{ABC} = 374^2 \times {}^{\cdot}0031526 - 374{\cdot}0000$
$= 66{\cdot}9731$

অর্থাৎ, $C_{ABC} = \sqrt{66{\cdot}9731/440{\cdot}9731}$
$= 0{\cdot}3897.$

সুতরাং $A$, $B$ এবং $C$ যৌথভাবে সংস্রবযুক্ত।

একত্রে $B$ ও $C$-র সঙ্গে $A$-র বহুল সংস্রব আছে কিনা দেখা যাক এরপর

এখানে $\chi^2_{A.BC} = 374(1{\cdot}02343 - 1)$
$= 8{\cdot}7628.$

## সারণী 9.5

| কোষ $(i, j, k)$ | কোষ-পরিসংখ্যা $f_{ijk}$ | $f_{i00} \times f_{0jk}$ | $(2)^2/(3)$ |
|---|---|---|---|
| (1) | (2) | (3) | (4) |
| $ABC$ | 130 | 45590 | ·37070 |
| $AB\gamma$ | 33 | 15040 | ·07241 |
| $A\beta C$ | 29 | 9165 | ·09176 |
| $A\beta\gamma$ | 43 | 18095 | ·10218 |
| $\alpha BC$ | 64 | 26966 | ·15189 |
| $\alpha\beta C$ | 31 | 8896 | ·10803 |
| $\alpha B\gamma$ | 10 | 5421 | ·01845 |
| $\alpha\beta\gamma$ | 34 | 10703 | ·10801 |
| মোট | 374 | — | 1·02343 |

সুতরাং $$C_{A.BC} = \sqrt{\frac{8\cdot7628}{374+8\cdot7628}} = 0\cdot1513$$

এবং $$T_{A.BC} = \sqrt{\frac{8\cdot7628}{374\sqrt{3}}} = 0\cdot1162$$

অর্থাৎ, একত্রে $B$ ও $C$-র ওপর $A$-র বহুল সংশ্রবের পরিমাণ তেমন উল্লেখযোগ্য নয়।

এখন $A$ ও $B$-র মধ্যে সামগ্রিক সংশ্রবের পরিমাণ

$$Q_{AB} = \frac{163 \times 44 - 72 \times 95}{163 \times 44 + 72 \times 95} = 0\cdot0237.$$

এটি কৃত্রিম কি না দেখার জন্য $C$-এর বিভিন্ন রূপের জন্য আলাদা আলাদা ভাবে $A$ ও $B$-র মধ্যে আংশিক সংশ্রবাঙ্কগুলির মান নির্ণয় করা যাক।

$$Q_{AB.C} = \frac{130 \times 10 - 29 \times 64}{130 \times 10 + 29 \times 64} = -0\cdot176.$$

$$Q_{AB.\gamma} = \frac{33 \times 34 - 43 \times 31}{33 \times 34 + 43 \times 31} = -0\cdot0859.$$

সুতরাং দেখা যাচ্ছে সামগ্রিক সংশ্রবাঙ্কের বিচারে $A$ ও $B$-র মধ্যে সামান্য ধনাত্মক সংশ্রব লক্ষিত হলেও প্রকৃতপক্ষে এটি ভ্রান্ত ধারণার উদ্রেক করে, কেননা আলাদা আলাদা ভাবে $C$-এর উপস্থিতিতে এবং অনুপস্থিতিতে $A$ ও $B$-র মধ্যে সংশ্রবের পরিমাণ ঋণাত্মক।

## 9.6 অনুশীলনী

9.1 গুণলক্ষণের যুগ্ম-বিভাজন বলতে কী বোঝ? প্রান্তিক বিভাজন ও সর্তাধীন বিভাজনের সংজ্ঞা দাও।

9.2 গুণলক্ষণের যুগ্ম-বিভাজনে কমপক্ষে কতগুলি পরিসংখ্যা দেওয়া থাকলে অন্যগুলি নির্ণয় করা সম্ভব? $2 \times 2 \times 2$ সারণীর ক্ষেত্রে চার-প্রস্থ পরিসংখ্যার উল্লেখ কর যেগুলির মধ্যে যে কোন একপ্রস্থ দেওয়া থাকলে সারণীটি সম্পূর্ণ জানা যায়।

নীচের উদাহরণে বিভিন্ন কোষ-পরিসংখ্যাগুলি নির্ণয় কর :

"একটি পরীক্ষায় মোট পরীক্ষার্থীর সংখ্যা 600, এদের মধ্যে বালিকাদের তুলনায় বালকেরা সংখ্যায় 16% বেশী। পরীক্ষায় উত্তীর্ণদের সংখ্যা অনুত্তীর্ণদের সংখ্যার চেয়ে 310 বেশী। বিজ্ঞান বিভাগে উত্তীর্ণ বালকদের সংখ্যা 300 এবং কলাবিভাগে অনুত্তীর্ণ বালিকাদের সংখ্যা 25। কলাবিভাগে মোট 135 জন পরীক্ষার্থীর মধ্যে অনুত্তীর্ণদের সংখ্যা 33, এবং পরীক্ষায় অকৃতকার্য বালকের সংখ্যা 18."

9.3 রাশিতথ্যের সামঞ্জস্য বলতে কী বোঝ? একাধিক গুণলক্ষণের যুগ্ম বিভাজন সংক্রান্ত একপ্রস্থ রাশিতথ্য সমঞ্জস হওয়ার সর্ত কী কী? একটি $2 \times 2 \times 2$ সারণীর ক্ষেত্রে এই সব সর্তের বীজগাণিতিক রূপগুলি দাও।

9.4 নীচের দুটি উদাহরণে প্রদত্ত তথ্যের মধ্যে কোন অসামঞ্জস্য আছে কি না বিচার কর :

(i) 57 জন মহিলাকে জিজ্ঞাসাবাদের পর নিম্নলিখিত তথ্যগুলি পাওয়া গেছে : গতমাসে এদের মধ্যে সিনেমা এবং থিয়েটার দুইই দেখেছেন 3 জন, সিনেমা দেখেছেন 39 জন এবং থিয়েটার দেখেছেন 22 জন।

(ii) $n = 1{,}000$ $\qquad f_{AB} = 143$

$f_A = 877$ $\qquad f_{AC} = 338$

$f_B = 286$ $\qquad f_{BC} = 135$

$f_C = 986$ $\qquad f_{ABC} = 107.$

9.5 বসন্ত রোগে আক্রান্ত কয়েকটি পরিবার সম্পর্কে সাম্প্রতিক একটি সমীক্ষায় জানা গেল, শতকরা 70 জন অধিবাসী এই রোগে আক্রান্ত হয়েছে এবং শতকরা 85 জনকে এই রোগের টীকা দেওয়া হয়েছিল। টীকা দেওয়া অধিবাসীদের মধ্যে কমপক্ষে শতকরা কতজন আক্রান্ত হয়েছে ?

9.6 দুটি গুণলক্ষণের যৌথ বিভাজনের ক্ষেত্রে সংস্রব এবং অনপেক্ষতার সংজ্ঞা দাও। আদর্শ সংস্রব মাপকের লক্ষণ কী কী ? কয়েকটি প্রচলিত সংস্রব মাপকের উল্লেখ কর এবং এদের সাধারণ ধর্মগুলি আলোচনা কর।

9.7 সংকেতচিহ্নগুলি প্রচলিত অর্থে গ্রহণ ক'রে প্রমাণ কর যে :

(i) $f^2_{AB} + f^2_{\alpha\beta} - f^2_{\alpha B} - f^2_{AB} = (f_A - f_\alpha)(f_B - f_\beta) + 2n\delta_{AB}.$

(ii) $$\delta_{AB} = \frac{f_B f_\beta}{n}\left\{\frac{f_{AB}}{f_B} - \frac{f_{A\beta}}{f_\beta}\right\}$$

$$= \frac{f_A f_\alpha}{n}\left\{\frac{f_{AB}}{f_A} - \frac{f_{\alpha B}}{f_\alpha}\right\}.$$

(iii) $$\frac{2Y_{AB}}{1 + Y^2_{AB}} = Q_{AB}.$$

9.8 সামগ্রিক, যৌথ, বহুল এবং আংশিক সংস্রবের সংজ্ঞা দাও 'কৃত্রিম সংস্রব' কী ? সংস্রব কৃত্রিম কিনা কীভাবে বোঝা যায় ?

9.9 নীচের সারণীতে প্রদত্ত রাশিতথ্য থেকে টীকাদান কলেরার প্রতিষেধক হিসাবে কতখানি কার্যকরী বিচার কর :

| | টীকা নিয়েছে | টীকা নেয়নি | মোট |
|---|---|---|---|
| কলেরায় আক্রান্ত | 37 | 459 | 496 |
| কলেরায় আক্রান্ত নয় | 191 | 1165 | 1356 |
| মোট | 228 | 1624 | 1852 |

9.10 পিতা ও পুত্রের উচ্চতা সংক্রান্ত নিম্নলিখিত তথ্য ধরা পড়েছে একটি সাম্প্রতিক সমীক্ষায় :

| | | পিতা | | | | |
|---|---|---|---|---|---|---|
| | | খুব লম্বা | লম্বা | মাঝারি | বেঁটে | মোট |
| পুত্র | খুব লম্বা | 30 | 20 | 20 | 2 | 72 |
| | লম্বা | 14 | 125 | 85 | 12 | 236 |
| | মাঝারি | 3 | 140 | 165 | 125 | 433 |
| | বেঁটে | 3 | 37 | 68 | 151 | 259 |
| | মোট | 50 | 322 | 338 | 290 | 1,000 |

পিতা ও পুত্রের উচ্চতার মধ্যে কতখানি সংশ্রব আছে পরীক্ষা কর।

9.11 একটি কারখানার শ্রমিকদের শারীরিক, মানসিক এবং স্নায়বিক স্বাস্থ্য সম্পর্কিত এক সমীক্ষায় নিম্নলিখিত তথ্য পাওয়া গেছে। লক্ষণ তিনটির মধ্যে বিভিন্ন ধরনের ( সামগ্রিক, যৌথ, বহুস এবং আংশিক ) সংশ্রবের পরিমাণ নির্ণয় কর।

| শারীরিক দৌর্বল্য | স্নায়বিক দৌর্বল্য | | | |
|---|---|---|---|---|
| | আছে | | নেই | |
| | জড়বুদ্ধি | জড়বুদ্ধি নয় | জড়বুদ্ধি | জড়বুদ্ধি নয় |
| আছে | 75 | 310 | 106 | 489 |
| নেই | 98 | 702 | 74 | 8415 |

## 9.7 নির্দেশিকা

1. Goon, A. M., Gupta, M. K. & Dasgupta, B. *Fundamentals of Statistics, Vol* 1. World Press, 1975.

2. Kendall, M. G. & Stuart, A. *Advanced Theory of Statistics Vol.* 1. Charles Griffin, 1960.

3. Wallis, W. A. & Roberts, H. V. *Statistics, a New Approach. Methuen*, 1957.

# 10 সহগতি ও নির্ভরণ : 1
## ( Correlation and Regression : 1 )

**10.1** পূর্ববর্তী একটি পরিচ্ছেদে একটি মাত্র চলের ভিত্তিতে কীভাবে পরিসংখ্যা-বিভাজন গঠন করতে হয় ও তার থেকে বিভাজনটির চারিত্রিক বৈশিষ্ট্য সম্পর্কে প্রয়োজনীয় তথ্য আহরণ করা যায় সে সম্পর্কে বিস্তারিত আলোচনা হয়েছে। এখন একই সঙ্গে দুটি বিভিন্ন চল সম্পর্কে রাশিতথ্য জানা থাকলে তার ভিত্তিতে তাদের পারস্পরিক সম্পর্ক সম্বন্ধে কীভাবে জ্ঞানলাভ করা যায়, তা আলোচনা ক'রে দেখা যাক। চল দুটির একটিকে $x$ ও অপরটিকে $y$ বলা হলে মনে করা যেতে পারে যে, (1) $x$ হচ্ছে কোন ব্যক্তির বয়স ও $y$ তার পিতার বয়স, (2) $x$ কোন ছাত্রের কলেজ পরীক্ষায় প্রাপ্ত শতকরা নম্বর ও $y$ তার বিশ্ববিদ্যালয় পরীক্ষায় প্রাপ্ত শতকরা নম্বর ; (3) $x$ কোন ব্যক্তির উচ্চতা ও $y$ তার ওজন ; (4) $y$ কোন উৎপন্ন ফসলের পরিমাণ ও $x$ ঐ ফসল উৎপাদনে ব্যবহৃত সারের পরিমাণ ; (5) $x$ হচ্ছে পাটের ওজন কাঁচা অবস্থায় ও $y$ ঐ পাটের ওজন শুকনো অবস্থায় ইত্যাদি। চলদুটির উভয়ে বিচ্ছিন্ন বা অবিচ্ছিন্ন অথবা একটি বিচ্ছিন্ন ও অপরটি অবিচ্ছিন্ন প্রকৃতির হতে আপত্তি নেই। এ ধরনের দ্বিচলভিত্তিক রাশিতথ্য থেকে প্রধানতঃ দুধরনের সমস্যা সমাধানের চেষ্টা করা হয়। এর একটি হ'ল এদের মধ্যে পারস্পরিক সম্পর্ক আছে কিনা ও থাকলে সেটি কী ধরনের ও কতখানি ঘনিষ্ঠ তা নির্ণয়ের চেষ্টা করা ; এবং অপরটি হ'ল এদের মধ্যে একটির সম্পর্কে কোন কিছু জানা থাকলে তার ভিত্তিতে অপরটি সম্পর্কে অনুমান করা অর্থাৎ প্রথমটির ওপর দ্বিতীয়টির এক ধরনের নির্ভরতা আবিষ্কারের চেষ্টা করে তার সাহায্যে প্রথমটিকে দ্বিতীয়টির সম্পর্কে অনুমান কাজে ব্যবহার করা। এজাতীয় প্রচেষ্টায় প্রাথমিক কাজ হচ্ছে লব্ধ উপাত্তকে সারণীতে প্রকাশ করা। এরকমের একটি সারণী নীচে দেওয়া হল ( সারণী 10.1 দ্রষ্টব্য )।

এই তথ্য সারণী থেকে আমরা মোটামুটিভাবে কলেজ ও বিশ্ববিদ্যালয়ের পরীক্ষায় প্রাপ্ত নম্বরের পারস্পরিক সম্পর্ক সম্বন্ধে কিছুটা ধারণা করতে পারি। অবশ্য এখানে 25 জন ছাত্রের নম্বর দেওয়া আছে। ফলে, মাত্র 25 জোড়া সংখ্যা অর্থাৎ $(x, y)$-এর 25টি মাত্র মান দেওয়া আছে। কিন্তু অনেক সময়

**সারণী 10.1**

**একটি কলেজ ও বিশ্ববিদ্যালয় পরীক্ষায় 25 জন ছাত্রের প্রাপ্ত শতকরা নম্বর।**

| ছাত্রের ক্রমিক সংখ্যা | কলেজ পরীক্ষায় প্রাপ্ত শতকরা নম্বর | বিশ্ববিদ্যালয় পরীক্ষায় প্রাপ্ত শতকরা নম্বর | ছাত্রের ক্রমিক সংখ্যা | কলেজ পরীক্ষায় প্রাপ্ত শতকরা নম্বর | বিশ্ববিদ্যালয় পরীক্ষায় প্রাপ্ত শতকরা নম্বর |
|---|---|---|---|---|---|
| (1) | (2) | (3) | (1) | (2) | (3) |
| 1 | 36 | 42 | 13 | 37 | 28 |
| 2 | 45 | 43 | 14 | 41 | 37 |
| 3 | 39 | 37 | 15 | 59 | 57 |
| 4 | 54 | 55 | 16 | 48 | 49 |
| 5 | 56 | 61 | 17 | 46 | 50 |
| 6 | 59 | 62 | 18 | 44 | 47 |
| 7 | 78 | 77 | 19 | 60 | 51 |
| 8 | 64 | 63 | 20 | 39 | 25 |
| 9 | 42 | 49 | 21 | 31 | 29 |
| 10 | 83 | 84 | 22 | 36 | 32 |
| 11 | 26 | 30 | 23 | 57 | 61 |
| 12 | 35 | 33 | 24 | 75 | 72 |
| | | | 25 | 61 | 68 |

উপাত্ত আরও বেশী সংখ্যক হতে পারে। যদি এরকম 500 জোড়া রাশি থাকে তবে এ ধরনের সারণি থেকে তথ্য নিষ্কাশন মুস্কিল হয়ে পড়বে। সেক্ষেত্রে আমরা দ্বিচলভিত্তিক পরিসংখ্যা-বিভাজন গঠনের কথা ভাবতে পারি। এ বিষয়ে তৃতীয় পরিচ্ছেদে বিস্তারিত আলোচনা হয়েছে। পরিসংখ্যা-বিভাজনটিকে একটি পরিসংখ্যা সারণীতে প্রকাশ করা হয়ে থাকে। নীচের পরিসংখ্যা সারণীটিতে (সারণী 10.2) 250 জন ছাত্র ইংরেজী ও অঙ্কে কোন পরীক্ষায় যে নম্বর (যথাক্রমে $y$ ও $x$) পেয়েছে তার ভিত্তিতে গঠিত একটি পরিসংখ্যা-বিভাজন দেখানো হয়েছে, যাতে ইংরেজী ও অঙ্কের জন্যে যথাক্রমে 8টি ও 10টি শ্রেণী

## সারণী 10.2

**250 জন ছাত্রের ইংরেজী ও অঙ্কে প্রাপ্ত নম্বরের ভিত্তিতে গঠিত দ্বিচল পরিসংখ্যা বিভাজন।**

ইংরেজীতে প্রাপ্ত শতকরা নম্বর ($y$)

| $x$ \ $y$ | $<$ 10·5 | 10·5 – 19·5 | 19·5 – 29·5 | 29·5 – 39·5 | 39·5 – 49·5 | 49·5 – 59·5 | 59·5 – 69·5 | 69·5 – 79·5 | মোট |
|---|---|---|---|---|---|---|---|---|---|
| $<$ 10·5 | 1 | 2 | 3 | 1 | | | | | 7 |
| 10·5 – 19·5 | 3 | 2 | 4 | 4 | 1 | | | | 14 |
| 19·5 – 29·5 | 1 | 3 | 8 | 5 | 1 | | | | 18 |
| 29·5 – 39·5 | | 1 | 15 | 9 | 2 | 1 | | | 28 |
| 39·5 – 49·5 | | | 12 | 21 | 11 | 4 | | | 48 |
| 49·5 – 59·5 | | 1 | 3 | 20 | 12 | 4 | 2 | | 42 |
| 59·5 – 69·5 | | | | 1 | 27 | 21 | 6 | | 55 |
| 69·5 – 79·5 | | | | | 25 | 3 | 1 | | 29 |
| 79·5 – 89·5 | | | | | | 4 | 1 | 1 | 6 |
| 89·5 – | | | | | | 1 | 1 | 1 | 3 |
| মোট | 5 | 9 | 45 | 61 | 79 | 38 | 11 | 2 | 250 |

ব্যবহার করা হয়েছে। ফলে, সারণীতে 80টি পৃথক পৃথক কোষ ব্যবহার করা হয়েছে। ঐ কোষগুলির অন্তর্বর্তী সংখ্যাগুলি ইংরেজী ও অঙ্কের এক এক জোড়া নম্বর নিয়ে গঠিত এক একটি শ্রেণীর পরিসংখ্যা নির্দেশ করছে। যেমন, ইংরেজী ($y$)-এর তৃতীয় শ্রেণী অর্থাৎ 19·5 – 29·5 শ্রেণী-অন্তর এবং অঙ্কের ($x$) পঞ্চম শ্রেণী অর্থাৎ 39·5 – 49·5 শ্রেণী-অন্তরের মধ্যে নম্বর পেয়েছে এমন ছাত্রের সংখ্যা হচ্ছে 12 অর্থাৎ $x$-এর পঞ্চম ও $y$-এর তৃতীয় অর্থাৎ (5, 3)-তম শ্রেণীর পরিসংখ্যা হচ্ছে 12. এখানে বলা হয় যে, (5, 3)-তম কোষের পরিসংখ্যা 12. প্রচলিত রীতি অনুযায়ী শায়ী পঙক্তিতে নির্দেশিত চলটির $i$-তম শ্রেণী ও প্রলম্ব পঙক্তিতে নির্দেশিত চলটির $j$-তম শ্রেণী—এই দুটি শ্রেণী যে কোষটিকে নির্দেশ করে তাকে $(i, j)$-তম কোষ বলে উল্লেখ করা হয়। ওপরের উদাহরণটিতে $i=1, 2, \cdots, 10$ এবং $j=1, 2, \cdots, 8$. সাধারণভাবে, যদি $i=1, 2, \ldots, k$ অর্থাৎ $x$ চলটির শ্রেণীসংখ্যা যদি $k$ হয় ও $j=1, 2, \ldots, l$ অর্থাৎ $y$ চলটির শ্রেণীসংখ্যা যদি $l$ হয়, তাহলে $k \times l$ সংখ্যক কোষগুলিতে প্রদর্শিত পরিসংখ্যাগুলি দেখাবে কীভাবে মোট পরিসংখ্যা $N$ (এক্ষেত্রে 250) $x$ ও $y$-এর $k \times l$ সংখ্যক কোষের মধ্যে বিভক্ত হয়েছে। সেজন্যে এই কোষগুলিতে প্রদর্শিত পরিসংখ্যাগুলি $N$ সংখ্যক মানের পরিসংখ্যা বিভাজন নির্দেশ করে বলে ধরা হয়। $(i, j)$-তম কোষের পরিসংখ্যাকে সাধারণতঃ $f_{ij}$ এই সংকেতচিহ্নের সাহায্যে প্রকাশ করা হয়। এখন, $\sum_j f_{ij} = f_{io}$,—এই সমষ্টিগুলি নির্ণয় করলে স্পষ্টই বোঝা যাবে যে, এই মানগুলি দেখাবে মোট পরিসংখ্যা $N$ কীভাবে কেবলমাত্র $x$ চলটির বিভিন্ন শ্রেণীর মধ্যে নিবেশিত রয়েছে। কাজেই $f_{io}$ $(i=1, 2, \ldots, k)$-এই সংখ্যাগুলি কেবলমাত্র একটি চল $x$-এর বিভাজন নির্দেশ করবে। ওপরের সারণীতে এদেরকে সর্বদক্ষিণস্থ প্রলম্ব পঙক্তিতে দেখানো হয়েছে। তেমনি, যদি $\sum_i f_{ij} = f_{oj}$ $(j=1, 2, \ldots, l)$ সংখ্যাগুলি নির্ণয় করা যায়, তবে তারা দেখায় মোট পরিসংখ্যা $N$ কীভাবে $y$ চলটির বিভিন্ন শ্রেণীগুলির মধ্যে ছড়ানো রয়েছে। কাজেই এরা কেবলমাত্র $y$ চলটির পরিসংখ্যা বিভাজন নির্দেশ করে। ওপরের সারণীতে এদেরকে সবচেয়ে নীচের শায়ী পঙক্তিতে দেখানো হয়েছে। $f_{io}$ ও $f_{oj}$ সংখ্যাগুলি সারণীটির প্রান্তীয় প্রলম্ব ও শায়ী পঙক্তিতে থাকে ব'লে এরা যে

দুটি বিভাজন নির্দেশ করে তাদেরকে প্রান্তীয় বিভাজন (marginal distribution) বলে ; $f_{io}$ ($f_{oj}$) পরিসংখ্যাগুলি $x(y)$-এর প্রান্তীয় পরিসংখ্যা-বিভাজন (marginal frequency distribution) নির্দেশ করে। ওপরের সারণীটিতে এ জাতীয় দু'প্রকারের বিভাজন [ যথা (1) সমস্ত কোষগুলির পরিসংখ্যা একত্রে চলদুটির যুগ্মবিভাজন নির্দেশ করে এবং (2) সারণীটির প্রান্তীয় পঙক্তিদ্বয় এক একটি চলের একক ও পৃথক প্রান্তীয় বিভাজন নির্দেশ করে ] ছাড়া আরও দু'শ্রেণীর বিভাজন নির্দিষ্ট রয়েছে। যে কোন একটি, মনে কর, $j$-তম প্রলম্ব পঙক্তিস্থিত শায়ী পঙক্তিগুলির খোপগুলিতে অবস্থিত পরিসংখ্যাগুলি স্বভাবতঃই $x$ চলের সর্তাধীন বিভাজন নির্দেশ করে, যার সর্তটি হচ্ছে এই যে, অপর চল $y$-এর মান ঐ $j$-তম শ্রেণীমধ্যে আবদ্ধ রয়েছে। $j$-এর প্রত্যেকটি মান 1, 2,..., $l$-এর জন্যেই এরকম এক একটি সর্তাধীন বিভাজন রয়েছে। এদের প্রত্যেকটিকেই এক একটি পঙক্তি বিভাজন বলা হয়। ঠিক তেমনি, $x$-এর $i$-তম শ্রেণীনির্দেশক শায়ী পঙক্তিটিকে স্থির রাখলে তৎস্থিত বিভিন্ন প্রলম্ব পঙক্তি অনুযায়ী বিভিন্ন কোষগুলির পরিসংখ্যা $y$ চলটির সর্তাধীন বিভাজন নির্দেশ করে যার আরোপিত সর্তটি হচ্ছে এই যে $x$-এর মান তার $i$-তম শ্রেণীগত মানগুলির মধ্যে সীমাবদ্ধ রয়েছে। $i$-এর প্রত্যেক মান 1, 2,..., $k$-এর জন্যে এমনি এক একটি সর্তাধীন বিভাজন রয়েছে। এদের প্রত্যেককেও এক একটি পঙক্তি বিভাজন (array distribution) বলে।

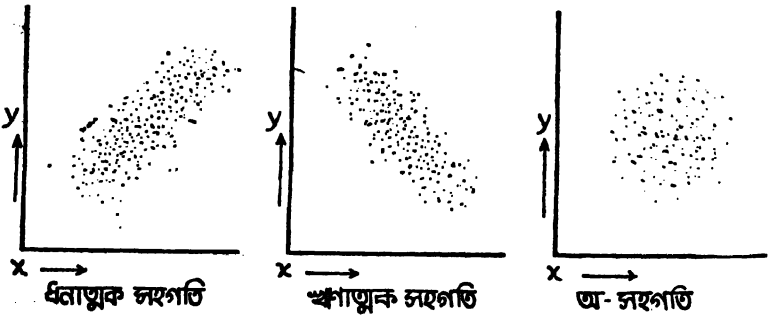
পূর্ববর্তী একটি পরিচ্ছেদে যেমন দেখা গেছে যে একচল বিভাজনের রাশি-তথ্যকে আয়তচিত্র, পরিসংখ্যা বহুভুজ ইত্যাদি চিত্রের মাধ্যমে প্রকাশ করা যায়, তেমনি দ্বিচলভিত্তিক রাশিতথ্যেরও চিত্রায়িত প্রকাশন সম্ভব। এজাতীয় সহজতম ও সর্বাধিক প্রচলিত চিত্রমাধ্যম হচ্ছে বিক্ষেপণ চিত্র (scatter diagram). এই চিত্র অঙ্কনের উদ্দেশ্যে $x$ ও $y$ চলদুটির জন্যে দুটি পরস্পর লম্ব অক্ষ নিয়ে $x$-কে অনুভূমিক ও $y$-কে উল্লম্ব অক্ষ বরাবর নেওয়া হয় এবং প্রত্যেক ব্যষ্টিকে তার জন্যে প্রাপ্ত $x$ ও $y$-এর মানদ্বয় ( ধর $x_0$ ও $y_0$) অনুযায়ী এক একটি বিন্দু ($x_0$, $y_0$) দ্বারা নির্দিষ্ট করা হয়, পূর্বোক্ত অক্ষদ্বয় অনুযায়ী যার ভুজ হচ্ছে $x_0$ ও কোটি হচ্ছে $y_0$. এক্ষেত্রে সুবিধামতো মূলবিন্দু ও মাপনা একক ব্যবহার করা হয়। যদি $n$ সংখ্যক ব্যষ্টি থাকে এবং $i$-তম ব্যষ্টির জন্যে $x$ ও $y$-এর মানদ্বয় যদি $x_i$ ও $y_i$ হয়, তবে লেখচিত্রটিতে এরকম $n$ সংখ্যক বিন্দু ($x_1$, $y_1$), ($x_2$, $y_2$), ..., ($x_i$, $y_i$), ..., ($x_n$, $y_n$) অঙ্কিত হবে। এই বিন্দুনিচয়ের গুচ্ছটিকেই বিক্ষেপণচিত্র

বলে। ব্যষ্টিসংখ্যা যদি খুব বেশী হয়, তাহলে বিক্ষেপণচিত্র খুব সার্থক ভূমিকা নিতে পারে না ; কারণ এক্ষেত্রে এটি খুব অর্থবাহী হবে না, এবং এর থেকে তথ্যনিষ্কাশন বাঞ্ছিতভাবে করা সম্ভব হবে না। এক্ষেত্রে অন্যধরনের চিত্র ব্যবহার করা যেতে পারে। এই প্রসঙ্গে উদাহরণস্বরূপ আয়তত‌লকের (stereogram) ব্যবহার তৃতীয় পরিচ্ছেদে কিছুটা আলোচিত হয়েছে।

## 10.2 সহগতি (Correlation) :

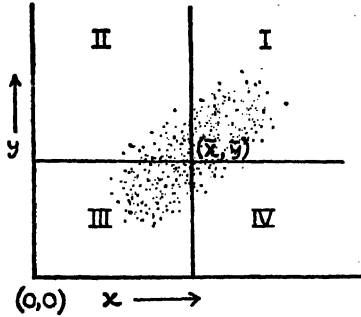
কোন দ্বিচলবিভাজন সম্পর্কে রাশিতথ্য হাতে থাকলে তার সাহায্যে চলদুটির পারস্পরিক সংস্রব বা ঘনিষ্ঠতা আছে কিনা, এবং থাকলে তা কী ধরনের এবং তার কী পরিমাণ, ইত্যাদি বিষয়ে কৌতূহল হওয়া স্বাভাবিক। এই কৌতূহল চরিতার্থ করতে যা সাহায্য করে তা হচ্ছে তাদের সহগতি। এই সহগতি বলতে কী বোঝায় দেখা যাক। সাধারণভাবে চলদুটির যেকোন একটির প্রত্যেক মানের জন্যে ( বা প্রত্যেক শ্রেণী-অন্তর মধ্যস্থ মানসমূহের জন্যে ) অপরটি যেকোন সংখ্যক মান গ্রহণ করতে পারে। মনে কর, এরকম একটি চল $x$-এর প্রত্যেক মানের জন্যে ( বা প্রত্যেক শ্রেণীভুক্ত মানগুচ্ছের জন্যে ) $y$ চলটি কয়েকটি মান নেয়। এখন, এই $y$ মানগুলির যৌগিক গড় নির্ণয় ক'রে যদি দেখা যায় যে, $x$-এর মান বৃদ্ধির সঙ্গে সঙ্গে ঐ গড়মানগুলিও সাধারণভাবে এবং গড়ে বেড়ে যেতে থাকে তবে বলব যে $x$ বাড়ার সঙ্গে সঙ্গে $y$-ও সাধারণতঃ বেড়ে যায় এবং এক্ষেত্রে বলা হয় যে, $x$ ও $y$-এর মধ্যে ধনাত্মক সহগতি রয়েছে। পক্ষান্তরে, $x$-এর মানবৃদ্ধির সঙ্গে সঙ্গে $y$-এর গড়মানগুলি যদি বেশীর ভাগ কমে যেতে থাকে তবে বলা হয় যে, $x$-এর বৃদ্ধির সঙ্গে সঙ্গে $y$ সাধারণতঃ কমতে থাকে এবং এক্ষেত্রে বলা হয় যে $x$ ও $y$-এর মধ্যে ঋণাত্মক সহগতি রয়েছে। এই দু'প্রকারের যেকোন একটি পরিস্থিতিতে বলা হয় যে, $x$ ও $y$ এর মধ্যে সহগতি আছে। আবার, যদি দেখা যায় যে $x$-এর বৃদ্ধির সঙ্গে সঙ্গে $y$-এর গড়মানগুলি মোটামুটি স্থির থাকে অথবা কয়েকটি কমে, কয়েকটি স্থির থাকে এবং কয়েকটি বাড়ে, কিন্তু একথা বলা যায় না যে তাদের বেশীর ভাগ বেড়েছে কি কমেছে অর্থাৎ তাদের গড় বৃদ্ধি বা হ্রাসের পরিমাণ লক্ষণীয় বৃদ্ধি বা হ্রাস কোন প্রবণতাই না দেখায়, তবে বলা হয় যে, $x$-এর হ্রাস-বৃদ্ধির সঙ্গে সঙ্গে $y$ গড়ে স্থির থাকে এবং এক্ষেত্রে বলা হয় যে, $x$ ও $y$-এর মধ্যে সহগতি নেই। এক্ষেত্রে $x$ ও $y$-কে পরস্পর সহগতিমুক্ত (uncorrelated) বলা হয়। এখানে একটি কথা বলা দরকার যে, এই সহগতির

সংজ্ঞায় একথাটি উহ্য রয়েছে যে, আমরা $x$ ও $y$-এর সেই ধরনের সংস্রবের কথাই মনে রেখেছি যাতে তাদের পারস্পরিক সম্পর্কটি অন্ততঃ মোটামুটিভাবে একটি ঋজুরৈখিক সূত্রে প্রকাশযোগ্য। কারণ, আমরা ধনাত্মক বা ঋণাত্মক সহগতি বলতে বুঝেছি একটির হ্রাস বা বৃদ্ধির সঙ্গে সঙ্গে অপরটির গড় হ্রাস বা বৃদ্ধির প্রবণতা অর্থাৎ একটির মান যখন একটি সরলরেখা ধরে এগিয়ে যাচ্ছে তখন অপর চলটির গড়মানগুলিরও মোটামুটিভাবে অপর একটি ঋজুরেখা ধরে এগিয়ে বা পেছিয়ে যাওয়ার ঝোঁক এবং সহগতিহীনতা বলতে বুঝেছি এই প্রবণতা প্রদর্শনের অভাব। সংক্ষেপে বলা যায় যে, পরস্পর সংস্রবযুক্ত দুটি চলের সহগতি হচ্ছে তাদের একটির পরিবর্তনে অপরটির ঋজুরৈখিক পরিবর্তনশীলতা।

এখন রাশিবিজ্ঞানসম্মত উপায়ে এই সহগতির পরিমাণ ও প্রকৃতি নির্ধারণ আমাদের উদ্দেশ্য। এ ব্যাপারে বিক্ষেপণচিত্র আমাদের খুব কাজে আসে। ওপরে সহগতির যে তিনপ্রকার পরিস্থিতির উল্লেখ করা হয়েছে তার প্রতিটি পরিস্থিতিতে বিক্ষেপণ চিত্রের চেহারা যে ধরনের হয়ে থাকে তা নীচের তিনটি ছবিতে দেখানো হ'ল।

চিত্র 10.1

ধনাত্মক, ঋণাত্মক এবং অ-সহগতি।

মনে করা যাক যে, দুটি চলের প্রত্যেকের $n$-সংখ্যক মান রয়েছে, যথাক্রমে $x_1, x_2, \cdots, x_n$ এবং $y_1, y_2, \cdots, y_n$; এদের মধ্যে $x_i$ ও $y_i$ হচ্ছে $i$-তম ব্যষ্টির জন্যে $x$ ও $y$-এর দুটি মান $(i=1, 2, \cdots, n)$. এখন, $\bar{x}=\frac{1}{n}\sum_{i=1}^{n} x_i$ ও $\bar{y}=\frac{1}{n}\sum_{i=1}^{n} y_i$ হচ্ছে চলদুটির গড়। এই তথ্যনির্ভর বিক্ষেপণচিত্রের মূলবিন্দু (0, 0) যদি $(\bar{x}, \bar{y})$

বিন্দুতে সরিয়ে নিয়ে প্রাথমিক $x$, $y$, ভুজকোটির সমান্তরাল ক'রে নতুন দুটি ভুজকোটি $x'=x-\bar{x}$, $y'=y-\bar{y}$ সম্বলিত একটি নতুন লেখ গঠন করা যায়, তাহলে $x'$, $y'$ ভুজকোটির ভিত্তিতে সমগ্র বিক্ষেপণচিত্রটি চারটি প্রকোষ্ঠে (quadrant) বিভক্ত হয়। তাহলে নতুন বিক্ষেপণচিত্রটির চেহারা নীচের চিত্রের মতো দাঁড়ায়।

চিত্র 10.2

বিক্ষেপণ চিত্র—চারিটি প্রকোষ্ঠ

মূল বিক্ষেপণচিত্রের $(x_i, y_i)$ বিন্দুগুলি এখন $(x'_i=x_i-\bar{x}, y'_i=y_i-\bar{y})$ বিন্দুরূপে নতুন $(x', y')$ ভুজকোটি সম্বলিত বিক্ষেপণচিত্রের চারটি প্রকোষ্ঠে ছড়ানো থাকবে। এখন, সাধারণ বোধশক্তিতে বোঝা যায় যে, যদি $x$ ও $y$-এর সহগতি ধনাত্মক হয়, তাহলে I ও III চিহ্নিত প্রকোষ্ঠের বিন্দুসংখ্যা অপর দুটি প্রকোষ্ঠের বিন্দুসংখ্যার চেয়ে বেশী হবে। আবার, I ও III চিহ্নিত প্রকোষ্ঠস্থিত বিন্দুগুলির ভুজকোটি $x'_i$ ও $y'_i$-গুলি সমচিহ্নবিশিষ্ট এবং II ও IV চিহ্নিত প্রকোষ্ঠস্থিত বিন্দুগুলির $x'_i$ ও $y'_i$ ভুজকোটিগুলি বিপরীত চিহ্নবিশিষ্ট হবে। এর ফলশ্রুতি হচ্ছে এই যে, $\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})=\sum_{i=1}^{n}x'_i\,y'_i$-এর মধ্যে ধনাত্মক পদসমূহের সমষ্টি ঋণাত্মক পদসমূহের চিহ্নবর্জিত সমষ্টির চেয়ে বেশী হবে। অর্থাৎ যদি $x$ ও $y$-এর সহগতি ধনাত্মক হয়, তবে $\sum_{i=1}^{n}x'_i\,y'_i>0$ হবে; পক্ষান্তরে সহগতি যদি ঋণাত্মক হয়, তবে অনুরূপ ভাবে দেখা যাবে যে $\sum_{i=1}^{n}x'_iy'_i<0$

হবে। ঠিক তেমনিভাবে বোঝা যায় যে, যদি $x$ ও $y$-এর সহগতি না থাকে তাহলে $\sum_{i=1}^{n} x'_i y'_i = 0$ হবে বা $\sum_{i=1}^{n} x'_i y'_i$-এর মান শূন্যের অন্ততঃ কাছাকাছি হবে। কাজেই, একথা বলা যাবে যে, যদি এমন হয় যে, $x$ ও $y$-এর মধ্যে ঋজুরৈখিক ধরনের সংশ্রব থাকে, তবে $\sum_{i=1}^{n} x'_i y'_i$ যথাক্রমে ধনাত্মক, ঋণাত্মক বা শূন্য হলে $x$ ও $y$-এর সহগতি ধনাত্মক, ঋণাত্মক বা অনুপস্থিত থাকবে। অবশ্য যদি $x$ ও $y$-এর কোন সংশ্রব থাকে, কিন্তু তা ঋজুরৈখিক ধরনের না হয়, তাহলে $\sum x'_i y'_i$-এর চিহ্ন বা মান থেকে তাদের প্রকৃত সহগতি বা সংশ্রব সম্পর্কে বিশেষ কিছু বলা যাবে না। এজন্যে সংশ্রব ঋজুরৈখিক হলে মনে করা যেতে পারে যে, $\sum x'_i y'_i = \sum (x_i - \bar{x})(y_i - \bar{y})$—এই অঙ্কটিকে $x$ ও $y$-এর সহগতির একটি উপযোগী মাপক (suitable measure) হিসেবে নেওয়া উচিত। এ বিষয়ে কিন্তু একটু অসুবিধে আছে। কারণ, $x$ ও $y$ চলদুটির মান $x_i$ ও $y_i$ সাধারণতঃ কোন বিশেষ এককের সূত্রে মাপা হবে ও তেমনিভাবেই প্রকাশিত হবে। যেমন, $x$ যদি কোন ব্যক্তির উচ্চতা ও $y$ যদি তার ওজন নির্দেশ করে, তবে $x_i, y_i$ মানগুলি সেন্টিমিটার এবং গ্রামের আকারে প্রকাশ করা যেতে পারে। এর ফল হবে এই যে, $\sum x'_i y'_i$ মাপকটিও কোন বিশেষ এককের মাধ্যমে প্রকাশিত হবে এবং $x'_i$ ও $y'_i$-এর একক যদি পরিবর্তিত হয়, তবে $\sum x'_i y'_i$-এর এককেরও পরিবর্তন হবে। কিন্তু সাধারণ বুদ্ধিতেই বোঝা যায় যে, সহগতির সাহায্যে $x$ ও $y$-এর যে সংশ্রব আমরা পরিমাপ করতে যাচ্ছি তার মাপনা এককের ওপর নির্ভর করা উচিত নয়। কাজেই, এখন আমাদের কর্তব্য হবে $\sum x'_i y'_i$-কে কোন উপায়ে মাপনা-একক নিরপেক্ষ করা। এর উপায় হচ্ছে $x'_i$ ও $y'_i$-এর উভয়কে $x$ ও $y$-এর প্রমাণ-বিচ্যুতি $s_x = \sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2}$ ও $s_y = \sqrt{\frac{1}{n} \sum (y_i - \bar{y})^2}$ দিয়ে ভাগ ক'রে নেওয়া। এরকম করার আরও সুবিধে এই যে এই সঙ্গে সহগতি সম্পর্কটিকে

$x$ ও $y$-এর বিস্তৃতি নিরপেক্ষও করা হয়ে যাবে। তাহলে আমরা যে মাপক পাই তা হচ্ছে $\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s_x}\right)\left(\frac{y_i-\bar{y}}{s_y}\right)$. কিন্তু একেও সহগতিমাপক হিসেবে ব্যবহার করার অন্তরায় হচ্ছে এই যে, এর মান মোট পদসংখ্যা বা মোট পরিসংখ্যা $n$-এর ওপর নির্ভরশীল। কিন্তু সহগতিমাপকের এই বাধ্যবাধকতা থাকা উচিত নয় কারণ এর ফলিতার্থ হবে এই যে, কেবল পরিসংখ্যা বৃদ্ধি বা হ্রাস ক'রে $x$ ও $y$-এর অন্তর্নিহিত সংস্রবের প্রকৃতি ও পরিমাণে হেরফের করা সম্ভব। কিন্তু বাস্তবিক পরিস্থিতি তা নয়। এই প্রতিবন্ধকের প্রতিকার পন্থা হচ্ছে $\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s_x}\right)\left(\frac{y_i-\bar{y}}{s_y}\right)$-কে সহগতির মাপক হিসেবে গ্রহণ করা। একে বলা হয় $x$ ও $y$-এর সহগাঙ্ক এবং $r_{xy}$ বা সংক্ষেপে $r$ সংকেতচিহ্ন দ্বারা একে প্রকাশ করা হয়। $r$ এর মান ধনাত্মক, ঋণাত্মক বা শূন্য হলে বুঝতে হবে $x$ ও $y$-এর সহগতিও যথাক্রমে ধনাত্মক, ঋণাত্মক বা অনুপস্থিত ; এবং এর বিপরীত ব্যাপারটিও সত্য অর্থাৎ সংস্রব ধনাত্মক, ঋণাত্মক বা অনুপস্থিত হলে $r$-এর মান যথাক্রমে ধনাত্মক, ঋণাত্মক বা শূন্য হবে।

সহগাঙ্ক $r=\dfrac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{s_x\, s_y}$ এর লব $\frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y})$-কে বলা হয় $x$ ও $y$-এর সহভেদমান (covariance). তাহলে, $x$ ও $y$-এর সহগাঙ্ককে আমরা লিখতে পারি

$$r_{xy}=r=\frac{x \text{ ও } y\text{-এর সহভেদমান}}{\sqrt{x \text{ এর ভেদমান}}\times\sqrt{y \text{ এর ভেদমান}}}$$

$$=\frac{x \text{ ও } y\text{-এর সহভেদমান}}{x\text{-এর প্রমাণ-বিচ্যুতি}\times y\text{-এর প্রমাণ-বিচ্যুতি}}$$

আমরা আরও লিখতে পারি

$$r=\frac{\frac{1}{n}\sum x_iy_i-\bar{x}\bar{y}}{\sqrt{\left(\frac{1}{n}\sum x_i^2-\bar{x}^2\right)\left(\frac{1}{n}\sum y_i^2-\bar{y}^2\right)}}$$

$$=\frac{n\,\Sigma x_iy_i-(\Sigma x_i)(\Sigma y_i)}{\sqrt{\{n\Sigma x_i^2-(\Sigma x_i)^2\}}\sqrt{\{n\Sigma y_i^2-(\Sigma y_i)^2\}}}.$$

যদি রাশিতথ্য অশ্রেণীবদ্ধ (ungrouped) রূপে থাকে, তবে সর্বশেষলিখিত সূত্রটিই $r$-এর মান নির্ণয়ের জন্যে ব্যবহারিক দিক থেকে সবচেয়ে উপযোগী।

**10·3. সহগাঙ্ক r-এর কয়েকটি ধর্ম :** (Some properties of the correlation coefficient $r$)

1. $r$ একটি বিশুদ্ধ সংখ্যা অর্থাৎ $x$ এবং $y$ যে এককের আকারে মাপা হয়েছে তার প্রভাব থেকে $r$ সম্পূর্ণ মুক্ত। মনে কর $x$ ও $y$ যথাক্রমে কিলোগ্রাম ও কিলোমিটারে প্রকাশিত সংখ্যা। কিন্তু $r$ হবে কেবল একটি প্রকৃত সংখ্যা; তার কোন মাপনা একক থাকবে না।

2. $-1 \leq r \leq 1$.

**প্রমাণ :** $$\frac{1}{n}\sum\left(\frac{x_i-\bar{x}}{s_x}\right)\left(\frac{y_i-\bar{y}}{s_y}\right)=\frac{1}{n}\sum u_i v_i\,;$$

[ $u_i = \dfrac{x_i-\bar{x}}{s_x}$ ও $v_i=\dfrac{y_i-\bar{y}}{s_y}$ লিখে ].

এখন, $\frac{1}{n}\sum\left(u_i+v_i\right)^2 \geq 0$, $u_i$ ও $v_i$-এর মান যে কোন প্রকৃত সংখ্যাই (real number) হোক না কেন। তাহলে,

$$\frac{1}{n}\sum u_i^2+\frac{1}{n}\sum v_i^2 \geq -\frac{2}{n}\sum u_i v_i. \qquad (10.1)$$

কিন্তু $\frac{1}{n}\sum u_i^2=\frac{1}{s_x^2}\,\frac{1}{n}\sum(x_i-\bar{x})^2=1$

এবং $\frac{1}{n}\sum v_i^2=\frac{1}{s_y^2}\,\frac{1}{n}\sum(y_i-\bar{y})^2=1.$

সুতরাং (10.1) থেকে পাই $2 \geq -2r$ অর্থাৎ $r \geq -1$. (10.2)

আবার, $\frac{1}{n}\sum\left(u_i-v_i\right)^2 \geq 0$, $u_i$ ও $v_i$-এর মান যে কোন প্রকৃত সংখ্যাই হোক না কেন।

অর্থাৎ $\frac{1}{n}\sum u_i^2+\frac{1}{n}\sum v_i^2 \geq \frac{2}{n}\sum u_i v_i$, অর্থাৎ $2 \geq 2r$

অর্থাৎ $r \leq 1$. ... ... (10.3)

সুতরাং (10.2) ও (10.3) থেকে পাওয়া যায়

$$-1 \leq r \leq 1. \qquad (10.4)$$

সহগাঙ্ক (coefficient of correlation or correlation coefficient) $r$ তার সর্বনিম্ন মান $-1$ গ্রহণ করে যখন প্রত্যেক $i=1, 2, \ldots, n$-এর জন্যে $u_i+v_i=0$ হয়, অর্থাৎ যখন $\frac{x_i-\bar{x}}{s_x}+\frac{y_i-\bar{y}}{s_y}=0$ অর্থাৎ যখন প্রত্যেক $i=1, 2, \ldots, n$-এর জন্যে $y_i=\bar{y}-\frac{s_y}{s_x}\left(x_i-\bar{x}\right)$ হয়। $\cdots$ (10.5)

পক্ষান্তরে, $r$ তার সর্বোচ্চ মান $+1$ গ্রহণ করে, যখন প্রত্যেক $i=1, 2, \ldots, n$ এর জন্যে $u_i-v_i=0$ অর্থাৎ $\frac{x_i-\bar{x}}{s_x}-\frac{y_i-\bar{y}}{s_y}=0$ হয়

অর্থাৎ যখন প্রতিটি $i=1, 2, \ldots, n$-এর জন্যে

$$y_i=\bar{y}+\frac{s_y}{s_x}(x_i-\bar{x}) \text{ হয়।} \qquad \cdots \quad (10.6)$$

এই উভয়ক্ষেত্রে $y$ এবং $x$-এর মধ্যে একটি যথাযথ ঋজুরৈখিক সম্পর্ক বিদ্যমান কারণ $x$ ও $y$-এর সম্পর্কটি $y=a+bx$ ধরনের একটি সমীকরণদ্বারা প্রকাশযোগ্য।

এখানে, $a=\bar{y}-\bar{x}\left(\pm\frac{s_y}{s_x}\right)$ এবং $b=\pm\frac{s_y}{s_x}$. এই উভয়ক্ষেত্রেই বলা হয় যে, $x$ ও $y$ এর মধ্যে সম্পূর্ণ সহগতি বা সম্পূর্ণ ঋজুরৈখিক সহগতি (linear correlation) রয়েছে। কারণ, এদের মধ্যে একটি অপরটির ঋজুরৈখিক অপেক্ষক। যখন, $r=-1$ হয় অর্থাৎ $b$ যখন $-\frac{s_y}{s_x}$, তখন বলা হয় যে, $x$ ও $y$ সম্পূর্ণ ঋণাত্মক সহগতিযুক্ত, এবং যখন $r=+1$ অর্থাৎ $b$ যখন $\frac{s_y}{s_x}$ তখন বলা হয় যে, $x$ ও $y$ সম্পূর্ণ ধনাত্মক সহগতিযুক্ত।

3. $x$ ও $y$-এর মূলবিন্দু এবং মাত্রার (origin and scale) পরিবর্তনে $r_{xy}$-এর মানের পরিমাণ পরিবর্তিত হয় না যদিও তার চিহ্নের পরিবর্তন হতে পারে। বিশদভাবে লেখা যায় যে, যদি $u=\frac{x-A}{C}$ এবং $v=\frac{y-B}{D}$ হয়, তাহলে রূপান্তরিত চল $u$ ও $v$-এর সহগাঙ্ক $r_{uv}$ ও $r_{xy}$-এর মধ্যে সম্পর্ক দাঁড়ায় $r_{uv}=\frac{|C|.|D|}{C.D}\, r_{xy}$.

**প্রমাণ ঃ** সংজ্ঞানুসারে,

$$r_{uv}=\frac{1}{s_u s_v}\cdot\frac{1}{n}\sum(u_i-\bar{u})(v_i-\bar{v}).$$

আমরা জানি যে, $u_i=\dfrac{x_i-A}{C}$, $v_i=\dfrac{y_i-B}{D}$ ; ফলে,

$$\bar{u}=\frac{\bar{x}-A}{C},\ \bar{v}=\frac{\bar{y}-B}{D},\ \left(u_i-\bar{u}\right)\left(v_i-\bar{v}\right)=\frac{1}{CD}(x_i-\bar{x})(y_i-\bar{y}),$$

$$s_u=\sqrt{\frac{1}{n}\sum\left(u_i-\bar{u}\right)^2}=\sqrt{\frac{1}{n}\cdot\sum\left(x_i-\bar{x}\right)\cdot\frac{1}{C^2}}$$

$$=\frac{1}{|C|}\sqrt{\frac{1}{n}\sum\left(x_i-\bar{x}\right)^2}=\frac{s_x}{|C|},$$

এবং $$s_v=\sqrt{\frac{1}{n}\sum\left(v_i-\bar{v}\right)^2}=\sqrt{\frac{1}{n}\sum\left(y_i-\bar{y}\right)^2\frac{1}{D^2}}$$

$$=\frac{1}{|D|}\sqrt{\frac{1}{n}\sum\left(y_i-\bar{y}\right)^2}=\frac{1}{|D|}s_y.$$

কাজেই $$r_{uv}=\frac{|C|.|D|}{C.D}\,\frac{1}{n}\sum\left(\frac{x_i-\bar{x}}{s_x}\right)\left(\frac{y_i-\bar{y}}{s_y}\right)=\frac{|C|.|D|}{C.D}\,r_{xy}.$$

যদি $C$ ও $D$ একই চিহ্নবিশিষ্ট হয়, তাহলে $r_{uv}=r_{xy}$ হবে, এবং $C$ ও $D$ যদি বিপরীত চিহ্নবিশিষ্ট হয়, তাহলে $r_{uv}=-r_{xy}$ হবে।

4. যদি $y=a+bx$ হয়, তবে

$r_{xy}=+1$, যদি $b>0$ হয়

$=-1$, যদি $b<0$ হয়।

এই ব্যাপারটি একটু আগেই ব্যাখ্যা করা হয়েছে।

## 10.4 গোষ্ঠী বা শ্রেণীবদ্ধ রাশিতথ্যের ভিত্তিতে সহগাঙ্ক নির্ণয় পদ্ধতি (Method of finding correlation coefficient from grouped data) ঃ

$x$ ও $y$ চলদুটির মানসংখ্যা $n$ যদি খুব বেশী হয়, তাহলে তাদের ওপর প্রাপ্ত রাশিতথ্যকে অনেক সময় একটি দ্বিচল পরিসংখ্যাসারণীর সাহায্যে লিপিবদ্ধ করা হয়। এ বিষয়ে আগেই উল্লেখ করা হয়েছে। এরকম সারণীভিত্তিক রাশিতথ্যের সাহায্যেও চলদুটির সহগতি পরিমাপ করা যায়। এ উদ্দেশ্যে নিম্নলিখিত প্রক্রিয়া অনুসরণ করা হয়।

মনে কর $x_i$ ও $y_j$ যথাক্রমে $x$ ও $y$ চলের $i$ ও $j$-তম শ্রেণীকুটির মধ্যবিন্দু এবং $f_{ij}$ হচ্ছে $(i, j)$-তম কোষের পরিসংখ্যা, $x$ ও $y$-এর শ্রেণীসংখ্যা যথাক্রমে $k$ ও $l$ এবং $\sum_{j=1}^{l} f_{ij}=f_{io}$, $\sum_{i=1}^{k} f_{ij}=f_{oj}$ ও মোট পরিসংখ্যা

$$\sum_{i=1}\sum_{j=1} f_{ij}=\sum_{i=1}^{k} f_{io}=\sum_{j=1}^{l} f_{oj}=n.$$

তাহলে, $x$ ও $y$-এর সহগাঙ্ক হিসেবে নেওয়া যায়

$$r_{xy}=r=\frac{\frac{1}{n}\sum\sum f_{ij}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\left\{\frac{1}{n}\sum f_{io}(x_i-\bar{x})^2\right\}\left\{\frac{1}{n}\sum f_{oj}(y_j-\bar{y})\right\}}}$$ -কে

এখানে, $\bar{x}=\frac{1}{n}\sum f_{io}x_i$ ও $\bar{y}=\frac{1}{n}\sum f_{oj}y_j$ হচ্ছে যথাক্রমে $x$ ও $y$-এর গড়। হিসেবের সুবিধার জন্যে $x$ ও $y$ -এর মূলবিন্দু ও মাপনামাত্রার পরিবর্তন করা যেতে পারে। তাহলে, $u=\frac{x-A}{C}$, $v=\frac{y-B}{D}$ লিখলে, এবং রীতি অনুযায়ী $A$ ও $B$-কে যথাক্রমে $x$ ও $y$-এর মাঝামাঝি কোন শ্রেণীর মধ্যবিন্দু এবং $C$ ও $D$-কে যথাক্রমে $x$ ও $y$-এর শ্রেণীদৈর্ঘ্য হিসেবে বেছে নিলে পাওয়া যাবে

$$r_{uv}=\frac{\frac{1}{n}\sum\sum f_{ij}(u_i-\bar{u})(v_j-\bar{v})}{\sqrt{\left\{\frac{1}{n}\sum f_{io}(u_i-\bar{u})^2\right\}\left\{\frac{1}{n}\sum f_{oj}(v_j-\bar{v})^2\right\}}}$$

$$=\frac{n\sum_i\sum_j f_{ij}u_iv_j-\left(\sum_i f_{io}u_i\right)\left(\sum_j f_{oj}v_j\right)\}}{\sqrt{\left\{n\sum_i f_{io}u_i^2-\left(\sum_i f_{io}u_i\right)^2\right\}\times\left\{n\sum_j f_{oj}v_j^2-\left(\sum f_{oj}v_j\right)^2\right\}}}.$$

এখানে $\bar{u} = \frac{1}{n} \sum_i f_{io} u_i = \frac{1}{n} \sum_i f_{io} \left( \frac{x_i - A}{C} \right)$

ও $\bar{v} = \frac{1}{n} \sum_j f_{oj} v_j = \frac{1}{n} \sum_j f_{oj} \left( \frac{y_j - B}{D} \right)$.

এছাড়া, $\sum_i \sum_j f_{ij} u_i v_j = \sum_i u_i \left( \sum_j f_{ij} v_j \right) = \sum_i u_i V_i$,

$V_i = \sum f_{ij} v_j$ লিখে

এবং $\sum \sum f_{ij} u_i v_j = \sum v_j \left( \sum f_{ij} u_i \right) = \sum_j v_j U_j$

$U_j = \sum_i f_{ij} u_i$ লিখে

কাজেই $\sum_i u_i V_i = \sum_j v_j U_j$. ... (10.7)

আবার, $\sum_i V_i = \sum_i \sum_j f_{ij} v_j = \sum_j v_j \left( \sum_i f_{ij} \right)$

$= \sum_j v_j f_{oj}$ ... (10.8)

এবং $\sum U_j = \sum \sum f_{ij} u_i = \sum u_i \left( \sum f_{ij} \right)$

$= \sum u_i f_{io}$. (10.9)

ওপরের এই ক'টি পারস্পরিক সম্পর্কের কথা মনে রাখলে রাশিতথ্যের দ্বিচল পরিসংখ্যা-বিভাজন থেকে সহগাঙ্ক সহজে নির্ণয় করার জন্যে নীচের সারণীটি ব্যবহার করা যায়। একে অনেক সময় সহগতি সারণী (correlation table) বলা হয়। পূর্বে আলোচিত 10.2 নং সারণীতে প্রদর্শিত দ্বিচল পরিসংখ্যা-বিভাজনের ভিত্তিতে এই সারণীটির বিভিন্ন কোষ এবং পঙ্‌ক্তিগুলি পূরণ করা হয়েছে। সারণীগঠনে হিসেবের শুদ্ধি পরীক্ষার (check) জন্যে (10.7)—(10.9) সম্পর্কগুলির সত্যতা লক্ষ্য ক'রে দেখে নেওয়া উচিত। এ উদ্দেশ্যে নীচে গঠিত

সারণীটিতে ( সারণী 10.3) তিনটি তীরচিহ্ন সাহায্যে বোঝানো হয়েছে যে এ সম্পর্ক তিনটি আলোচ্য রাশিতথ্যের ক্ষেত্রে খাটে।

সারণি 10.2-এ প্রদত্ত রাশিতথ্যের ভিত্তিতে ইংরেজী ও অঙ্কে প্রাপ্ত নম্বরের সহগাঙ্ক নির্ণয় করতে গিয়ে লেখা যাক

$x_i = i$-তম শ্রেণীর মধ্যবিন্দু ( $x$-এর অর্থাৎ অঙ্কের নম্বরের জন্যে )

$y_j = j$-তম শ্রেণীর মধ্যবিন্দু ( $y$-এর অর্থাৎ ইংরেজীর নম্বরের জন্যে )

$$u_i = \frac{x_i - A}{C}, \quad A = 45, \ C = 10,$$

$$v_j = \frac{y_j - B}{D}, \quad B = 45, \ D = 10.$$

$$n = \text{মোট ছাত্রসংখ্যা} = 250.$$

তাহলে, সহগতি সারণীটি নিম্নরূপ দাঁড়ায়।

তাহলে, সংজ্ঞানুযায়ী, অঙ্ক ও ইংরেজীতে প্রাপ্ত নম্বরের সহগাঙ্ক হচ্ছে

$$r = \frac{n \sum u_i V_i - \left(\sum u_i f_{io}\right)\left(\sum v_j f_{oj}\right)}{\sqrt{\left\{n \sum u_i^2 f_{io} - \left(\sum u_i f_{io}\right)^2\right\}\left\{n \sum v_j^2 f_{oj} - \left(\sum v_j f_{oj}\right)^2\right\}}}$$

$$= \frac{250 \times 523 - (-106)(-132)}{\sqrt{250 \times 994 - (-106)^2}\sqrt{250 \times 502 - (-132)^2}}$$

$$= \frac{116758}{\sqrt{(237264 \times 108076)}}.$$

তাই $\log_{10} r = \bar{1}.862806$ এবং $r = \cdot 729$.

## 10.5 ঔপপত্তিক বা তত্ত্বগত দ্বিচল বিভাজন (Theoretical bivariate distribution) :

এতক্ষণ সহগতির আলোচনা যে ক্ষেত্রে $X$ ও $Y$ চলদুটির কেবলমাত্র সসীম-সংখ্যক মান রয়েছে তাতেই সীমাবদ্ধ রয়েছে। আসলে কিন্তু প্রায়শঃই $X$ ও $Y$-এর সসীম $n$ সংখ্যক প্রদত্ত মানকে একটি অজ্ঞাত পূর্ণক থেকে গৃহীত নমুনা ব'লে গণ্য করা এবং পূর্ণকটিতে এমন আরও অনেক মান রয়েছে ব'লে স্বীকার করা উচিত হবে। এই পূর্ণকটিকে কোন তত্ত্বগত বিভাজন দ্বারা সূচিত করা যেতে পারে এবং সেই তত্ত্বগত বিভাজনটি দুটি অবিচ্ছিন্ন বা বিচ্ছিন্ন চলের যুগ্মবিভাজন

## সারণী 10.3

## সহগতি সারণী

## 250 জন ছাত্রের ইংরেজী ও অঙ্কের নম্বরের সহগতি নির্ণয়

$y$ ( ইংরেজীর শতকরা নম্বর ) →

$x$ ( অঙ্কের শতকরা নম্বর ) →

| শ্রেণী মধ্যক $y_j$ / শ্রেণী মধ্যক $x_i$ | | 5 | 15 | 25 | 35 | 45 | 55 | 65 | 75 | $f_{i0}$ | $u_if_{i0}$ | $u_i^2f_{i0}$ | $V_i$ | $u_iV_i$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | $v_j$ / $u_i$ | −4 | −3 | −2 | −1 | 0 | 1 | 2 | 3 | | | | | |
| 5 | −5 | 1 | 2 | 3 | 1 | | | | | 7 | −35 | 175 | −17 | 85 |
| 15 | −4 | 3 | 2 | 4 | 4 | 1 | | | | 14 | −56 | 224 | −30 | 120 |
| 25 | −3 | 1 | 3 | 8 | 5 | 1 | | | | 18 | −54 | 162 | −34 | 102 |
| 35 | −2 | | 1 | 15 | 9 | 2 | 1 | | | 28 | −56 | 112 | −41 | 82 |
| 45 | −1 | | | 12 | 21 | 11 | 4 | | | 48 | −48 | 48 | −41 | 41 |
| 55 | 0 | | 1 | 3 | 20 | 12 | 4 | 2 | | 42 | 0 | 0 | −21 | 0 |
| 65 | 1 | | | | 1 | 27 | 21 | 6 | | 55 | 55 | 55 | 32 | 32 |
| 75 | 2 | | | | | 25 | 3 | 1 | | 29 | 58 | 116 | 5 | 10 |
| 85 | 3 | | | | | | 4 | 1 | 1 | 6 | 18 | 54 | 9 | 27 |
| 95 | 4 | | | | | | 1 | 1 | 1 | 3 | 12 | 48 | 6 | 24 |
| $f_{0j}$ | | 5 | 9 | 45 | 61 | 79 | 38 | 11 | 2 | 250 | −106 | 994 | −132 | 523 |
| $v_jf_{0j}$ | | −20 | −27 | −90 | −61 | 0 | 38 | 22 | 6 | −132 | | | | |
| $v_j^2f_{0j}$ | | 80 | 81 | 180 | 61 | 0 | 38 | 44 | 18 | 502 | | | | |
| $U_j$ | | −20 | −29 | −97 | −74 | 55 | 37 | 15 | 7 | −106 | | | | |
| $v_jU_j$ | | 80 | 87 | 194 | 74 | 0 | 37 | 30 | 21 | 523 | | | | |

হতে পারে। এক্ষেত্রে ঐ পূর্ণক বা তত্ত্বগত বিভাজনের ভিত্তিতে সহগাঙ্ক $\rho$-কে $\dfrac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$ হিসেবে নেওয়া হবে। এখানে cov $(X, Y)$, $\sigma_X$ ও $\sigma_Y$ হচ্ছে $X$ ও $Y$ চলদুটির সহভেদমান এবং যথাক্রমে তাদের প্রমাণ-বিচ্যুতি। বিচ্ছিন্ন বা অবিচ্ছিন্ন সম্ভাবনা চলের ক্ষেত্রে cov $(X, Y)$, $\sigma_X$ ও $\sigma_Y$-এর সংজ্ঞা সপ্তম পরিচ্ছেদে আলোচিত হয়েছে।

সহগাঙ্কের সংজ্ঞা থেকে স্পষ্টই প্রতীয়মান হচ্ছে যে, চলদুটি সম্ভাবনাতত্ত্বানুযায়ী নির্ভরতাশূন্য হলে $\rho$-এর মান শূন্য হবে অর্থাৎ তাদের সহগতি অনুপস্থিত থাকবে। এর বিপরীত প্রতিজ্ঞাটি অবশ্য সর্বদা সিদ্ধ নাও হতে পারে। কোন কোন ক্ষেত্রে যদি $\rho = 0$ হয় অর্থাৎ $X$ ও $Y$-এর যদি সহগতি না থাকে তাহলেও তারা সম্ভাবনাতত্ত্বানুযায়ী পরস্পর নির্ভরশীল হতে পারে। এই ব্যাপারটি একটু বিস্তারিতভাবে আলোচনা করে দেখা যাক।

প্রথম ক্ষেত্র : চলদুটি বিচ্ছিন্ন।

ধরা যাক, $X$ ও $Y$ দুটি বিচ্ছিন্ন সম্ভাবনা চল এবং তাদের মানগুলি যথাক্রমে $x_1, x_2, \ldots, x_i, \ldots$ এবং $y_1, y_2, \ldots y_j, \ldots$ এবং তাদের যুগ্মবিভাজনটি একটি অপেক্ষক $P$-এর সাহায্যে এরূপে নির্দেশিত যে,

$P_{ij} = P[X = x_i, Y = y_j]$ অর্থাৎ $X$ ও $Y$-এর যুগপৎ যথাক্রমে $x_i$ ও $y_j$ মান গ্রহণ করার সম্ভাবনা হচ্ছে $P_{ij}$. তাদের প্রান্তীয় বিভাজন-দুটিকে যথাক্রমে

$$p_i = \sum P_{ij} = \sum P\Big[X = x_i,\ Y = y_j\Big] = P\Big[X = x_i\Big]$$

ও $$q_j = \sum_i P_{ij} = \sum P\Big[X = x_i,\ Y = y_j\Big] = P\Big[Y = y_j\Big]$$ দ্বারা এবং তাদের গাণিতিক প্রত্যাশাকে যথাক্রমে $\mu_X$ ও $\mu_Y$ দ্বারা নির্দেশ করা হলে যদি $X$ ও $Y$ সম্ভাবনাতত্ত্বানুযায়ী পরস্পর অনধীন হয়, তবে

সব $i, j = 1, 2, \ldots$ এর জন্যে

$$P_{ij} = p_i q_j$$

এবং $$\text{cov}(X, Y) = \sum \sum_j \Big(x_i - \mu_X\Big)\Big(y_j - \mu_Y\Big) p_i q_j$$

$$= \Big\{\sum \Big(x_i - \mu_X\Big) p_i\Big\}\Big\{\sum \Big(y_j - \mu_Y\Big) q_j\Big\} = 0.$$

কাজেই $\rho=0$ অর্থাৎ চলদুটি পরস্পর সহগতিমুক্ত।

দ্বিতীয় ক্ষেত্র : চলদুটি অবিচ্ছিন্ন এবং সম্ভাবনা-ঘনত্ব-অপেক্ষক সম্বলিত। অবিচ্ছিন্ন চল $X$ ও $Y$-এর যুগ্মবিভাজনের সম্ভাবনা-ঘনত্ব-অপেক্ষককে $f(x, y)$ এবং তাদের প্রান্তীয় বিভাজনের সম্ভাবনা-ঘনত্ব-অপেক্ষককে $g(x)$ এবং $h(y)$ দ্বারা এবং তাদের গাণিতিক প্রত্যাশাকে $\mu_X$ ও $\mu_Y$ দ্বারা সূচিত করা হলে যদি তারা সম্ভাবনাতত্ত্বানুযায়ী পরস্পর অনধীন হয়, তবে প্রত্যেক $x$ ও $y$-এর জন্যে

$f(x, y)=g(x)\,h(y)$ এবং চলদুটির মানসীমা যথাক্রমে $(\alpha, \beta)$ ও $(\gamma, \delta)$ হলে

$$\text{cov }(X, Y)=\int_{\alpha}^{\beta}\int_{\gamma}^{\delta}(x-\mu_X)(y-\mu_Y)g(x)h(y)dx\,dy$$
$$=\left\{\int_{\alpha}^{\beta}(x-\mu_X)g(x)dx\right\}\left\{\int_{\gamma}^{\delta}(y-\mu_Y)h(y)\,dy\right\}=0$$

এবং তার ফলে $\rho=0$ হবে অর্থাৎ তারা সহগতিমুক্ত হবে।

পক্ষান্তরে, মনে করা যাক $X$ ও $Y$ সম্ভাবনা চলদুটির মধ্যে $X^2+Y^2=k^2$—এই গাণিতিক সম্পর্কটি রয়েছে। ধরা যাক যে $X$ চলটি কেবলমাত্র $\pm i$ $(i=0, 1, 2,\ldots, k)$ মানগুলি ধারণ করে এবং $P[X=\pm i]=\frac{1}{2k+1}$. বলা বাহুল্য, $Y$ সেই সমস্ত মান ধারণ করবে যেগুলি $X^2+Y^2=k^2$ এই সম্পর্কসূত্রের সঙ্গে সামঞ্জস্যপূর্ণ। তাহলে $Y$-এর মানগুলি হচ্ছে

$$y_j=\sqrt{k^2-j^2},\ j=0,\ \pm i\ (i=1, 2,\ldots, k)$$

এবং $P[X=\pm i, Y=y_j]=0$, যদি $j\neq i$ হয়,

$=P[X=\pm i]$, যদি $j=i$ হয়,

$$=\frac{1}{2k+1},\ i=0,\ \pm 1,\ \pm 2,\ldots,\ \pm k.$$

সুতরাং $E(X)=\frac{1}{2k+1}\sum_i i=0$, কারণ, $i=0,\ \pm 1,\ \pm 2,\ldots,\ \pm k.$

$E(XY)=\frac{1}{(2k+1)}\sum_i i\sqrt{k^2-i^2}=0$, কারণ এখানে

$$i=0,\ \pm 1,\ \pm 2,\ldots,\ \pm k.$$

সুতরাং $\text{cov }(X, Y)=E(XY)-E(X)E(Y)=0.$

কাজেই $\rho=0$ অর্থাৎ $X$ ও $Y$ পরস্পর সহগতিমুক্ত।

অথচ $X$ ও $Y$ যদিও সহগতিশূন্য, তারা মোটেই সংশ্রবহীন নয়, কারণ স্পষ্টতঃই তারা একটি সুস্পষ্ট গাণিতিক সম্পর্কযুক্ত এবং $Y$ এর প্রত্যেকটি মানই $X$-এর গৃহীত মানের ওপর প্রত্যক্ষভাবে নির্ভরশীল। আরও স্পষ্টভাবে দেখা যায় যে,

$$P[X=i,\ Y=\sqrt{k^2-i^2}] \neq P[X=i] \times P[Y=\sqrt{k^2-i^2}]$$

কারণ, $P[X=i,\ Y=\sqrt{k^2-i^2}]=P[X=i]$

এবং $P[Y=\sqrt{k^2-i^2}] \neq 1.$

কাজেই চল দুটি পরস্পর অনধীন নয়।

এখন, আরও একটি উদাহরণ নেওয়া যাক। মনে করা যাক, দুটি বিচ্ছিন্ন সম্ভাবনা চল $X$ ও $Y$-এর যুগ্ম বিভাজন নিম্নলিখিতরূপ। $X$ ও $Y$-এর মানগুলি মনে করা যাক $x_1=-1,\ x_2=0,\ x_3=+1$

এবং $y_1=-1,\ y_2=0$ ও $y_3=+1.$

## সারণী 10.4

### $X$ ও $Y$-এর যুগ্ম সম্ভাবনা বিভাজন

| $Y$ \ $X$ | $-1$ | $0$ | $1$ | প্রান্তীয় সমষ্টি |
|---|---|---|---|---|
| $-1$ | 0 | $\frac{1}{4}$ | 0 | $\frac{1}{4}$ |
| $0$ | $\frac{1}{4}$ | 0 | $\frac{1}{4}$ | $\frac{1}{2}$ |
| $1$ | 0 | $\frac{1}{4}$ | 0 | $\frac{1}{4}$ |
| প্রান্তীয় সমষ্টি | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{4}$ | 1. |

এই সারণীতে $(i-j)$ তম কোষের সংখ্যা নির্দেশ করছে $P[X=x_i,\ Y=y_j]$ $(i, j=1, 2, 3)$-এর মান।

তাহলে, $E(X)=-1\times\frac{1}{4}+0\times\frac{1}{2}+1\times\frac{1}{4}=0$

$E(Y)=-1\times\frac{1}{4}+0\times\frac{1}{2}+1\times\frac{1}{4}=0.$

$$E(XY)=(-1\times\overline{1}\times 0)+(-1\times 0\times \tfrac{1}{2})+(-1\times 1\times 0)+(0\times -1\times \tfrac{1}{2})+(0\times 0\times 0)+(0\times 1\times \tfrac{1}{2})+(1\times \overline{1}\times 0)+(1\times 0\times \tfrac{1}{2})+(1\times 1\times 0)=0.$$

কাজেই, $\text{cov}(X, Y)=E(XY)-E(X)E(Y)=0.$

কিন্তু, $P[X=0, Y=0]=0$, অথচ $P[X=0]=\frac{1}{2}$ ও $P[Y=0]=\frac{1}{2}$.

ফলে, $P[X=0, Y=0]\neq P[X=0].P[Y=0]$.

কাজেই চলদুটি সম্ভাবনাতত্ত্বানুযায়ী পরস্পর অনধীন নয়।

অনুশীলনীতে একটি উদাহরণ [10.1 দ্রষ্টব্য ] থেকে দেখা যাবে যে, কোন কোন ক্ষেত্রে $\rho$-এর মান শূন্য হলেই চলদুটি সম্ভাবনাতত্ত্বানুযায়ী নির্ভরতাশূন্য হবে।

## 10.6 নির্ভরণ তত্ত্ব (Theory of Regression) :

নির্ভরণ তত্ত্বের গভীর ও ব্যাপক আলোচনার অবকাশ এ গ্রন্থে নেই। আমরা এর তাৎপর্যটুকু অল্প কথায় ব'লে এর ব্যবহারিক দিকটি একটু বিস্তারিত ভাবে বলার চেষ্টা করব।

অনেক সময় আমাদের আলোচ্য দুটি চলের মধ্যে একটি অপরটির ওপর কোন না কোনভাবে নির্ভরশীল ব'লে মনে করার কারণ ঘটে। যেমন, কোন ফসলের উৎপাদনের পরিমাণ তার চাষে ব্যবহৃত সারের পরিমাণের ওপর নির্ভর করে; কোন সমাজে ব্যক্তিবর্গের ব্যয়ের পরিমাণ তাদের আয়ের পরিমাণের ওপর নির্ভর করে, ইত্যাদি। এ সমস্ত ক্ষেত্রে দুটি চলের ভূমিকার মধ্যে যে একটি পার্থক্য রয়েছে সেকথা প্রথমেই স্বীকার ক'রে নেওয়া দরকার। পার্থক্যটি এই যে একটির ওপর অপরটি নির্ভরশীল। এই নির্ভরশীলতার কথা স্মরণে রেখে, যে চলটি অপরটির ওপর নির্ভর করছে সেটিকে নির্ভরী চল (dependent variable) ও অপরটিকে স্বনির্ভর বা অনপেক্ষ বা অনধীন চল (independent variable) বলে উল্লেখ করা হয়। এরপর স্বভাবতঃই দেখবার চেষ্টা করা হয় এই নির্ভরতার পরিপ্রেক্ষিতে স্বনির্ভর চলটির সম্পর্কে কোন জ্ঞাত তথ্য থেকে অপর চলটি সম্পর্কে বিজ্ঞানসম্মত ও নির্ভরযোগ্যভাবে কোন অজ্ঞাততথ্য অনুমান করা যায় কিনা এবং করা হলে তার মূল্যায়ন সম্ভব কিনা; অর্থাৎ একটির কোন মানের জন্যে অপরটির কী মান হওয়া উচিত সে সম্পর্কে ভবিষ্যদ্বাণী করা বা পূর্বাভাষ দেওয়া যায় কিনা তা দেখা আমাদের মুখ্য উদ্দেশ্য। এখন, মনে কর

$X$ হচ্ছে একটি অনপেক্ষ চল ও $Y$ তার ওপরে নির্ভরশীল অপর একটি চল। $X$-এর প্রত্যেক মানের জন্যে সাধারণভাবে $Y$-এর কতগুলি মান থাকে এবং তারা এক একটি স্তবক ( বা গুচ্ছ ) রচনা করে (array). এরকম প্রত্যেক স্তবকের $Y$ মানগুলি এক একটি সর্তাধীন বিভাজন গঠন করে ব'লে ধরা যেতে পারে। এখানে সর্তটি হচ্ছে এই যে এই সমস্ত $Y$ মানের জন্যে $X$-এর মান কোন একটি সংখ্যা $x_i$-তে স্থিরীকৃত। এদেরকে স্তবক বিভাজন বা পঙ্‌ক্তি বিভাজন (array distribution) বলা যেতে পারে। এখন $X$-এর উপর $Y$-এর নির্ভরতা বিচার করার একটি উপায় হচ্ছে $X$-এর মানের পরিবর্তনের সঙ্গে সঙ্গে এই স্তবক বিভাজনগুলির পরিবর্তনের রীতিটি অনুসরণ করা। এই উদ্দেশ্যে সমগ্র স্তবক বিভাজনটির কথা চিন্তা না ক'রে, দেখবার চেষ্টা করা হয় কীভাবে স্তবক বিভাজনের গড়গুলি $X$-এর সঙ্গে সঙ্গে পরিবর্তিত হয়। এই সম্পর্ক অনুধাবনের জন্যে প্রকৃষ্ট পন্থা হচ্ছে স্তবক গড়গুলির (array means) সঙ্গে $X$ চলটির কোন গাণিতিক সম্পর্ক স্থাপনের চেষ্টা করা। $X$-এর কোন নির্দিষ্ট মান $x$-এর জন্যে স্তবক গড়টিকে $E(Y/x)$ দ্বারা চিহ্নিত করা যেতে পারে। এটি হচ্ছে $X$ এর মান $x$-এ আবদ্ধ থাকার সর্তে $Y$ চলের গাণিতিক প্রত্যাশা। এখন, যদি $\psi$ এমন একটি অপেক্ষক হয় যার জন্যে $E(Y/x)=\psi(x)$, তাহলে $\psi$-কে বলা হয় $X$-এর ওপর $Y$-এর নির্ভরণ অপেক্ষক (Regression Function). এখন একটি লেখচিত্রে $x$-কে ভুজ ও $\psi(x)$-কে কোটি বরাবর ধ'রে $(x, \psi(x))$, বিন্দুগুলি স্থাপন করলে তাদের ওপর দিয়ে যে রেখা অতিক্রম করবে তাকে বলা হবে $X$-এর ওপর $Y$-এর নির্ভরণ রেখা। $\psi(x)$ যদি $x$-এর একটি ঋজুরৈখিক অপেক্ষক হয় অর্থাৎ যদি $\psi(x)=a+bx$ লেখা যায়, তাহলে এই নির্ভরণ রেখাটি একটি সরলরেখা হবে এবং এক্ষেত্রে বলা হবে যে, $X$ এর ওপর $Y$ এর নির্ভরণ হচ্ছে ঋজুরৈখিক। প্রকৃত নির্ভরণের স্বরূপ সাধারণতঃ জানা যায় না। $X$ চলের ওপর $Y$ চলের নির্ভরণের স্বরূপ জানতে হলে তাদের কয়েকটি মানকে উপযুক্তরূপে বিশ্লেষণ করা ছাড়া উপায় নেই। কাজেই নির্ভরণের স্বরূপটি প্রকৃতপক্ষে ঠিক কী ধরনের তা জানা সম্ভব নয়। কিন্তু প্রদত্ত রাশিতথ্যের সাহায্যে সে সম্পর্কে অনুমান করার চেষ্টা করা যেতে পারে। এই উদ্দেশ্যে অনেক সময় ধরা হয় যে, নির্ভরণটি ঋজুরৈখিক ধরনের; বাস্তবিক, প্রকৃত নির্ভরণটি যাই হোক না কেন $Y_x=\alpha+\beta x$ এই রেখাটিকে প্রকৃত নির্ভরণ অপেক্ষক $\psi(x)$-এর একটি আসন্ন রূপ হিসেবে ধরা যেতে পারে; অর্থাৎ $X$-এর কোন মান $x$-এর জন্যে $Y$-এর মানের প্রাক্‌কলক হিসেবে

$Y_x$-কে ধরা হয়। $Y_x$ মানগুলি স্পষ্টতঃই একটি ঋজুরেখার ওপর থাকে। একে বলা হয় $X$-এর ওপর $Y$-এর নির্ভরণের সাযুজ্যরক্ষাকারী সরলরেখা (fitted line of regression). এখন, লক্ষণীয় হচ্ছে যে, $Y_x = \alpha + \beta x$ রেখার সমীকরণে $\alpha$ ও $\beta$ ছুটি অজ্ঞাতরাশি। কাজেই প্রদত্ত নমুনালব্ধ রাশিতথ্যের ভিত্তিতে এদের ছুটি প্রাক্কলক নির্ণয় ক'রে নেওয়া দরকার। এই প্রাক্কলক নির্ণয়ে যে নীতি অনুসরণ করা হয় তাকে বলে লঘিষ্ঠ বর্গনীতি (principle of least squares) ও এই প্রাক্কলন পদ্ধতিকে বলে লঘিষ্ঠ বর্গ পদ্ধতি। এই নীতি ও পদ্ধতিটি একটু ব্যাখ্যা করা যাক।

মনে কর $X$-এর মান যখন $x_i$, তখন $Y$ চলটি $n_i$ সংখ্যক বিভিন্ন মান গ্রহণ ক'রে এবং সেগুলি হচ্ছে $y_{i1}, \ldots\ldots, y_{in_i}$. একটি লেখচিত্রের ভুজ ও কোটি বরাবর যথাক্রমে $x_i$ ও $y_{ij}$-কে ($j=1, 2, \cdots, n_i$ ; $i=1, 2, \cdots, k$) নিয়ে $(x_i, y_{ij})$ বিন্দুগুলি সাধারণতঃ অঙ্কিত হয়ে থাকে। এখন, $\bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$ হচ্ছে যে স্তবকে $X$-এর মান $x_i$, সেই স্তবকের জন্মে $Y$-এর গড়। তাহলে, $\bar{y}_i$-এর সঙ্গে $x_i$-এর সম্পর্ক নির্ণায়ক রেখাটিই হচ্ছে $X$-এর ওপর $Y$-এর প্রকৃত নমুনাভিত্তিক নির্ভরণ। কিন্তু সসীম সংখ্যক $(x_i, \bar{y}_i)$-এর মানের ভিত্তিতে এদের মধ্যে কোন গাণিতিক সম্পর্ক-সূত্র প্রতিষ্ঠা করা সম্ভব নয়। কাজেই $(x_i, \bar{y}_i)$ বিন্দুসমূহ সংযোগকারী রেখার সঙ্গে সাযুজ্যরক্ষাকারী হিসেবে $Y_x = \alpha + \beta x$ ধরনের ($\alpha$, $\beta$ অজ্ঞাত) একটি সরলরেখা এখন নির্ণয় করার চেষ্টা করা হয়, যার সাহায্যে $X$-এর ওপর $Y$-এর নির্ভরণের একটি ঋজুরৈখিক অনুমাপক পাওয়া যেতে পারে। স্বভাবতঃই এটা বাঞ্ছনীয় যে $\alpha$ ও $\beta$ যেন এমনভাবে নির্ধারিত হয় যাতে $(x_i, Y_i = \alpha + \beta\, x_i)$ বিন্দুগুলি $(x_i, y_{ij})$ বিন্দুগুলির যথাসম্ভব কাছাকাছি থাকে। কিন্তু প্রত্যেকটি $(x_i, Y_i)$ বিন্দুতো আর প্রত্যেকটি $(x_i, y_{ij})$ বিন্দুর সমীপবর্তী হতে পারে না। তাই চেষ্টা করা হয় যাতে $(x_i, Y_i)$ ও $(x_i, y_{ij})$ বিন্দুগুলির দূরত্ব সামগ্রিকভাবে যথাসম্ভব কম হয়। এইজন্মে চেষ্টা করা হয় $\alpha$ ও $\beta$ যেন এমনভাবে বেছে নেওয়া হয় যাতে

$S = \sum_i \sum_j (y_{ij} - \alpha - \beta x_i)^2$-এর মান সর্বাপেক্ষা কম হয়। এই হ'ল লঘিষ্ঠ বর্গনীতি। এই নীতিপ্রয়োগের ফলশ্রুতি হচ্ছে এই যে, এরূপে নির্ধারিত

রেখাটির ধর্ম হবে এই যে, $(x_i, Y_i)$ ও $(x_i, y_{ij})$ বিন্দুগুলির দূরত্বের বর্গগুলির সমষ্টি সবচেয়ে কম হবে। এখানে দূরত্ব বলতে অবশ্য লম্ব দূরত্ব নয়। আসলে বিন্দুগুলির দূরত্ব মাপা হবে যে লেখচিত্রে $(x_i, y_{ij})$ বিন্দুগুলি সন্নিবিষ্ট হয়েছে তার উল্লম্ব অক্ষ বরাবর।

এখন, $S=\sum_i \sum_j (y_{ij}-a-\beta x_i)^2$-কে $a$ ও $\beta$ এর উপযুক্ত মান নির্বাচনের সাহায্যে লঘিষ্ঠ করার জন্যে আমরা অন্তর্কলন পদ্ধতির সাহায্য নেব এবং $a$ ও $\beta$-কে

$$0=\frac{\partial S}{\partial a}=\frac{\partial}{\partial a}\sum_i \sum_j (y_{ij}-a-\beta x_i)^2$$

$$=-2\sum_i \sum_j (y_{ij}-a-\beta x_i) \quad \cdots \quad (10.10)$$

এবং $$0=\frac{\partial S}{\partial \beta}=\frac{\partial}{\partial \beta}\sum_i \sum_j (y_{ij}-a-\beta x_i)^2$$

$$=-2\sum' \sum_j' x_i(y_{ij}-a-\beta x_i) \quad \cdots \quad (10.11)$$

এই সমীকরণ-দুটির মূল হিসেবে সমাধান ক'রে নির্ণয় করা হবে। (10.10) ও (10'11) থেকে যথাক্রমে পাওয়া যায়

$$\sum_i \sum_j y_{ij}=na+\beta\sum n_i x_i, \quad \cdots \quad (10.12)$$

$$\left(n=\sum n_i \text{ লিখে}\right)$$

এবং $$\sum_i \sum_j y_{ij}x_i=a\sum n_i x_i+\beta \sum n_i x_i^2. \quad \cdots \quad (10.13)$$

এই সমীকরণ-দুটিকে [ (10.12) ও (10.13)-কে এবং (10.10) ও (10.11)-কে ] বলে নর্ম্যাল বা মৌল সমীকরণ (normal equations). এদের সমাধান বের করতে গিয়ে পাওয়া যাবে

$$a=\bar{y}-\beta\bar{x}, \left[\text{ কারণ } \bar{y}=\frac{1}{n}\sum_i \sum_j y_{ij}=\frac{1}{n}\sum_i n_i\bar{y}_i \right.$$

$$\left. \text{ও } \bar{x}=\frac{1}{n}\sum n_i x_i \right]$$

এবং $$\sum_i \sum_j y_{ij}x_i = \bar{y}\sum n_i x_i - \beta\bar{x}\sum n_i x_i + \beta\sum n_i x_i^2$$

অর্থাৎ $$\beta = \frac{\sum_i \sum_j y_{ij}x_i - \bar{y}\sum n_i x_i}{\sum_i n_i x_i^2 - n\bar{x}^2} = \frac{\sum\sum (y_{ij} - \bar{y})(x_i - \bar{x})}{\sum n_i (x_i - \bar{x})^2}$$

$$\frac{\sum_i n_i(\bar{y}_i - \bar{y})(x_i - \bar{x})}{\sum n_i(x_i - \bar{x})^2} = \hat{\beta} \qquad (10.14)$$

এবং $$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}. \qquad (10.15)$$

এখানে $\hat{\alpha}$ ও $\hat{\beta}$ বোঝাচ্ছে লঘিষ্ঠ বর্গনীতি অনুযায়ী নির্ধারিত অজ্ঞাতরাশি $\alpha$ ও $\beta$-এর প্রাক্কলক। $S = \sum_i \sum_j (y_{ij} - \alpha - \beta x_i)^2$-এর মানকে লঘিষ্ঠ রেখে $\alpha$ ও $\beta$-এর মান এমনভাবে নির্ণয় করার পদ্ধতিকে লঘিষ্ঠ বর্গপদ্ধতি বলে। এখানে $y_{ij}$ হচ্ছে $Y$ চলের অবেক্ষিত মান (observed value), $Y_i = \alpha + \beta x_i$ হচ্ছে প্রত্যেক $x_i$-এর জন্যে $Y$ চলের অনুমিত মান এবং $(y_{ij} - Y_i)$ হচ্ছে যে সমস্ত ব্যষ্টির জন্যে $X$-এর মান $x_i$-তে নিবদ্ধ তাদের $Y$ চলের মানগুলির মধ্যে একটি $y_{ij}$ থেকে $Y_i$-এর পার্থক্য। একে পরিভাষানুযায়ী বলা হয় অবশিষ্টাংশ (residual). কারণ, আসল $y_{ij}$-এর কিছুটা অংশ নির্ধারিত ঋজুরৈখিক নির্ভরণ-অপেক্ষক $Y_x = \alpha + \beta x$-এর মাধ্যমে নির্ণীত বা ব্যাখ্যাত হয়েছে $Y_i = \alpha + \beta x_i$ দ্বারা। কিন্তু এই নির্ভরণ অপেক্ষক $y_{ij}$-এর সবটুকু নির্দেশ করতে পারছে না এবং $(y_{ij} - Y_i)$ অংশটুকু অবশিষ্ট রয়ে গেছে। লঘিষ্ঠ বর্গপদ্ধতিতে নির্ভরণ ঋজু রেখাটি (regression line) এমনভাবে নির্ণয় করতে হয় যেন এইজাতীয় সবকটি অবশিষ্টাংশের বর্গের সমষ্টি ক্ষুদ্রতম হয়।

যদি প্রত্যেক স্তবকে একটি ক'রে মাত্র মান থাকে অর্থাৎ যদি প্রত্যেক $i = 1, 2, \ldots, n$-এর জন্যে $n_i = 1$ হয়, তবে (10.14) থেকে দেখা যায় যে,

$$\frac{\sum \left(y_i - \bar{y}\right)\left(x_i - \bar{x}\right)}{\sum \left(x_i - \bar{x}\right)^2} = \frac{\text{cov}(X, Y)}{V(X)} = \frac{r s_x s_y}{s_x^2}$$

অর্থাৎ $\hat{\beta} = r\frac{s_y}{s_x}$ … (10.16)

[এখানে $s_x = \sqrt{V(X)} = \sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2}$

ও $s_y = \sqrt{V(Y)} = \sqrt{\frac{1}{n}\sum(y_i - \bar{y})^2}$.]

অবশ্য সাধারণভাবেও (10.16) এ উল্লিখিত সম্পর্কটি সত্য।

সারল্যের অনুরোধে আমরা এখন থেকে $\hat{\alpha}$-কে $a$ ও $\hat{\beta}$-কে $b$ অথবা $byx$ সংকেত সাহায্যে প্রকাশ করব। তাহলে দাঁড়ালো

$$b = r\frac{s_y}{s_x} \text{ ও } a = \bar{y} - b\bar{x} = \bar{y} - r\frac{s_y}{s_x}\bar{x}$$

এবং অনুমিত নির্ভরণ সরলরেখাটির সমীকরণ হচ্ছে

$$\hat{Y}x = a + bx = \bar{y} + b(x - \bar{x}) = \bar{y} + r\frac{s_y}{s_x}(x - \bar{x}).$$

এখানে $b$-কে বলা হয় $X$-এর ওপর $Y$-এর নির্ভরণাঙ্ক (coefficient of regression). একটি লেখচিত্রে যদি $Y_x = \bar{y} + r\frac{s_y}{s_x}(x - \bar{x})$ রেখাটি অঙ্কিত হয়, তাহলে $b = r\frac{s_y}{s_x}$ নির্দেশ করবে অনুভূমিক রেখার (abscissa) ওপর নির্ভরণ-রেখাটির নতির পরিমাণ (inclination). যদি নতিকোণটির (angle of inclination) পরিমাণ $\theta$ হয়, তবে $\tan\theta = b = r\frac{s_y}{s_x}$ এবং $a = \bar{y} - r\frac{s_y}{s_x}$ নির্দেশ করবে ভুজকোটির মূলবিন্দু থেকে এই রেখা ও উল্লম্ব অক্ষের ছেদবিন্দুর উল্লম্ব অক্ষ বরাবর দূরত্ব।

ওপরে যে তথ্যসাহায্যে $X$-এর ওপর $Y$-এর নির্ভরণ নির্ণীত হল, সেই তথ্যের ভিত্তিতেই চলদুটির ভূমিকা পরিবর্তন ক'রে $Y$-এর ওপর $X$-এর নির্ভরণও নির্ণয় করা যায়। সেক্ষেত্রে নির্ভরণরেখাটির সমীকরণ হবে $X_y = \gamma + \delta y$ ধরনের এবং লঘিষ্ঠ বর্গপদ্ধতি অনুসারে এর অনুমিত সমীকরণ হবে

$$\hat{X}_y = \hat{\gamma} + \hat{\delta}y \text{ এবং এতে } \hat{\gamma} = \bar{x} - \hat{\delta}\bar{y} \text{ ও}$$

$$\hat{\delta} = r\frac{s_x}{s_y} = b_{xy} \text{ ( ধরা যাক ) দাঁড়াবে। অর্থাৎ}$$

$\hat{X}_y = \bar{x} + r\frac{s_x}{s_y}(y-\bar{y})$ এবং $r\frac{s_x}{s_y}$ হবে $Y$-এর ওপর $X$-এর নির্ভরণাঙ্ক।
এখন, লেখচিত্রে যদি

$\hat{X}_y = \bar{x} + r\frac{s_x}{s_y}(y-\bar{y})$ এবং $\hat{Y}_x = \bar{y} + r\frac{s_y}{s_x}(x-\bar{x})$ ঋজুরেখা-দুটি আঁকা যায়, তাহলে তাদের ছেদবিন্দু হবে $(\bar{x}, \bar{y})$. ছেদবিন্দুতে রেখাদুটির অন্তর্ভূত সূক্ষ্মকোণটি যদি $\omega$ হয় এবং যথাক্রমে $\theta$ ও $\phi$ যদি অনুভূমিক অক্ষে $\hat{Y}_x = \hat{\alpha} + \hat{\beta}x$ ও $\hat{X}_y = \hat{\gamma} + \hat{\delta}y$ রেখাদুটির নতিকোণের পরিমাণ হয়, তাহলে

$$\tan\phi = r\frac{s_x}{s_y},\ \tan\theta = r\frac{s_y}{s_x} \text{ এবং}$$

$$\tan\omega = \tan(\phi-\theta) = \frac{\tan\phi - \tan\theta}{1+\tan\phi\tan\theta} = \frac{\dfrac{s_y}{rs_x} - r\dfrac{s_y}{s_x}}{1+\dfrac{s_y}{rs_x}\times\dfrac{rs_y}{s_x}}$$

$$= \frac{1-r^2}{r}\times\frac{s_x s_y}{s_x^2+s_y^2}.$$

**টীকা:** লক্ষণীয় যে, $\hat{\beta}$ ও $\hat{\delta}$ সমচিহ্নযুক্ত।

## 10.7 নির্ভরণরেখা সংক্রান্ত কয়েকটি তথ্য: (Some facts about regression lines)

(1) মনে কর $U = \dfrac{X-A}{C}$ ও $V = \dfrac{Y-B}{D}$.

এবং $u = \dfrac{x-A}{C}$ ও $v = \dfrac{y-B}{D}$.

তাহলে, $U$-এর ওপর $V$-এর নির্ভরণাঙ্ককে $b_{vu}$ লিখলে,

$$b_{vu} = \frac{\text{cov}(V,U)}{V(u)} = \frac{\text{cov}(U,V)}{s_u^2}$$

[ $V(U) = s_u^2$ ও $V(V) = s_v^2$ লিখে ]

আবার $b_{yx} = \dfrac{\text{cov}(X,Y)}{V(X)} = \dfrac{\text{cov}(X,Y)}{s_x^2}$.

কিন্তু $\text{cov}(U,V) = \dfrac{1}{n}\sum\left(u-\bar{u}\right)\left(v-\bar{v}\right) = \dfrac{1}{CD}\dfrac{1}{n}\sum(x-\bar{x})(y-\bar{y})$

$$= \frac{1}{CD}\operatorname{cov}(X, Y) \text{ এবং } V(y) = \frac{1}{n}\sum (u-\bar{u})^2$$

$$= \frac{1}{C^2}\cdot\frac{1}{n}\sum (x-\bar{x})^2 = \frac{s_x^2}{C^2}.$$

সুতরাং $$b_{vu} = \frac{\dfrac{\operatorname{cov}(X, Y)}{CD}}{\dfrac{s_x^2}{C^2}} = \frac{C}{D}\,\frac{\operatorname{cov}(X, Y)}{s_x^2} = \frac{C}{D}b_{yx} \quad \cdots \quad (10.17)$$

অর্থাৎ $b_{yx} = \dfrac{D}{C}b_{vu}$. লক্ষণীয় যে, $\bar{y} = B + D\bar{v}$ ও $\bar{x} = A + C\bar{u}$.

এছাড়া, $V_u = \bar{v} + r_{uv}\dfrac{s_v}{s_u}(u-\bar{u})$ হচ্ছে $U$-এর ওপর $V$-এর নির্ভরণরেখার সমীকরণ। ফলে,

$$Y_x = \bar{y} + r_{xy}\frac{s_y}{s_x}(x-\bar{x}) = B + Dv + \frac{D}{C}b_{vu}(x - A - C\bar{u}). \quad \cdots \quad (10.18)$$

(10.17) ও (10.18) সম্পর্ক-দুটির সার্থকতা হচ্ছে এই যে, $Y$ ও $X$ এর মূলবিন্দু ও মাপনামাত্রা পরিবর্তন ক'রে $V$ ও $U$-এর গড়, প্রমাণ-বিচ্যুতি ও তাদের সহগাঙ্ক ও $U$-এর ওপর $V$-এর নির্ভরণাঙ্ক নির্ণয় করলে তাদের মাধ্যমেই $X$-এর ওপর $Y$-এর নির্ভরণাঙ্ক ও নির্ভরণরেখার সমীকরণ সহজেই নির্ণয় করা যায়।

(2) আমরা পেয়েছি $\hat{Y}_i = \hat{a} + \hat{\beta}x_i$, $\hat{a} = \bar{y} - \hat{\beta}\bar{x}$, $\hat{\beta} = r\dfrac{s_y}{s_x}$, এবং এইগুলি পাওয়া গেছে ( সারল্যের অনুরোধে প্রত্যেক $i = 1,\ldots,n$-এর জন্যে $n_i = 1$ ধরে )

$$\sum (y_i - a - \beta x_i) = 0 \text{ অর্থাৎ } \sum y_i = na + \beta\sum x_i \quad (10.19)$$

এবং $$\sum x_i(y_i - a - \beta x_i) = 0$$

অর্থাৎ $$\sum x_i y_i = a\sum x_i + \beta\sum x_i^2 \quad \cdots \quad (10.20)$$

এই দুটি নর্ম্যাল সমীকরণ ব্যবহার ক'রে। তাহলে,

$$\bar{\hat{Y}} = \frac{1}{n}\sum_{i=1}^{n}\hat{Y}_i = \frac{1}{n}\sum\{\bar{y} + \hat{\beta}(x_i - \bar{x})\} = \bar{y} + \hat{\beta}\frac{1}{n}\sum (x_i - \bar{x}) = \bar{y}.$$

অর্থাৎ অনুমিত রেখা থেকে নির্ণীত মানগুলির গড় এবং $Y$ চলের অবেক্ষিত মানগুলির গড় উভয়েই সমান।

এছাড়া, $e_i = y_i - \hat{Y}_i = y_i - \hat{\alpha} - \hat{\beta} x_i$ হচ্ছে
নির্ভরণ সাযুজ্যরেখা প্রদত্ত মানের অবশিষ্টাংশ (residual). এর থেকে পাই

$$\sum e_i = \sum (y_i - \hat{Y}_i) = \sum (y_i - \alpha - \beta x_i) = 0 \qquad (10.21)$$

অর্থাৎ $\quad - \quad \frac{1}{n} \sum e_i.$

অর্থাৎ ব্যষ্টিগতভাবে অবশিষ্টাংশগুলির মান যাই হোক, তাদের সমষ্টি হচ্ছে সর্বদাই শূন্য।

(3) $$V(\hat{Y}) = \frac{1}{n} \sum_i (\hat{Y}_i - \hat{Y})^2 = \frac{1}{n} \sum_i \{\bar{y} + \hat{\beta}(x_i - \bar{x}) - \bar{y}\}^2$$

$$= \hat{\beta}^2 \frac{1}{n} \sum (x_i - \bar{x})^2 = r^2 \frac{s_y^2}{s_x^2} s_x^2 = r^2 s_y^2 = r^2 V(Y).$$

সুতরাং $\quad r^2 = \frac{V(\hat{Y})}{V(Y)}. \qquad (10.22)$

সুতরাং $\quad |r| = \frac{\hat{Y}\text{-এর প্রমাণ-বিচ্যুতি}}{Y\text{-এর প্রমাণ-বিচ্যুতি}}$।

কাজেই, $|r|$ হচ্ছে $Y$-এর যে অংশ অনুমিত নির্ভরণরেখা থেকে নির্ণীত হয়েছে তার প্রমাণ-বিচ্যুতি এবং $Y$ এর মোট প্রমাণ-বিচ্যুতির অনুপাত।

যেহেতু, $\quad -1 \leq r \leq 1$, অর্থাৎ $r^2 \leq 1$, এটা স্পষ্টই দেখা যাচ্ছে যে,

$$V(\hat{Y}) = r^2 V(Y) \leq V(Y).$$

(4) $$s_{y \cdot x}^2 = V(e) = \frac{1}{n} \sum (e_i - \bar{e})^2 = \frac{1}{n} \sum e_i^2$$

$$= \frac{1}{n} \sum (y_i - \hat{Y}_i)^2 = \frac{1}{n} \sum \{(y_i - \bar{y}) - \hat{\beta}(x_i - \bar{x})\}^2$$

$$= \frac{1}{n} \sum (y_i - \bar{y})^2 + \hat{\beta}^2 \frac{1}{n} \sum (x_i - \bar{x})^2 - 2\hat{\beta} \frac{1}{n} \sum (y_i - \bar{y})(x_i - \bar{x})$$

$$= \frac{1}{n} \left[ \sum (y_i - \bar{y})^2 - \hat{\beta}^2 \sum (x_i - \bar{x})^2 \right] = s_y^2 (1 - r^2)$$

এবং $e$-এর প্রমাণ-বিচ্যুতি হচ্ছে

$$s_{y \cdot x} = s_y \sqrt{1 - r^2}$$

লক্ষণীয় যে, $V(e) > 0$ কারণ এটি একটি ভেদমান। ফলে, $s_y^2(1-r^2) > 0$ অর্থাৎ $1-r^2 > 0$, অর্থাৎ $r^2 < 1$.

সুতরাং $-1 < r < 1$—এই ফলটির এটি একটি বিকল্প প্রমাণ।

যদি $r = \pm 1$ হয় তবে $V(e) = 0$ ও ফলে প্রত্যেক $i = 1, \ldots, n$-এর জন্যে $e_i = \bar{e} = 0$ অর্থাৎ $y_i = \hat{Y}_i$ হবে। সেক্ষেত্রে বিক্ষেপণ চিত্রে প্রত্যেক $(x_i, y_i)$ বিন্দুই একটি সরলরেখার ওপর থাকবে অর্থাৎ অনুমিত নির্ভরণরেখা থেকে $Y$-এর প্রতিটি মান ভ্রান্তিশূন্যভাবে নির্ণয় করা যাবে। এক্ষেত্রে নির্ভরণরেখাটিকে $Y$-এর আদর্শ প্রাক্কলক সূত্র হিসেবে গণ্য করা যাবে, কারণ $X$-এর প্রতিটি মান $x$-এর জন্যে $Y$-এর অনুমিত মান $Y_x = \hat{\alpha} + \hat{\beta}x$ এবং তদনুযায়ী $Y$-এর আসল মান অভিন্ন হয়ে যাবে। এস্থলে, যদি নমুনাটিই পূর্ণক হয়, তবে $X$-এর ওপর $Y$-এর প্রকৃত নির্ভরণটিই ঋজুরৈখিক হবে। কিন্তু যেহেতু সাধারণতঃ নমুনাটি পূর্ণকের অংশমাত্র একথা জোর ক'রে বলা যাবে না যে, পূর্ণকেও এই নির্ভরণ ঋজুরৈখিক হবেই। তবে, $r = \pm 1$ হলে $X$-এর ওপর $Y$-এর নমুনালব্ধ নির্ভরণসূত্র যথার্থ ঋজুরৈখিক হওয়ার ফলে এটা আশা করা খুবই সঙ্গত হবে যে সমগ্র পূর্ণকটিতেও নির্ভরণ খুব সম্ভবতঃ ঋজুরৈখিক প্রকৃতিসম্পন্ন।

পক্ষান্তরে যদি $r = 0$ হয়, তাহলে, $V(e) = s_y^2$ হবে অর্থাৎ নির্ভরণরেখা-সাহায্যে অনুমিত মানগুলির অবশিষ্টাংশগুলির ভেদমান আর মূল $y$-গুলির ভেদমান সমান হয়ে যাবে। কাজেই এখানে নির্ণীত নির্ভরণরেখাটি $y$-এর অনুমাপক হিসেবে ব্যবহারের অযোগ্য হবে। এটা খুব স্পষ্টই বোঝা যাচ্ছে আরও এই কারণে যে, এক্ষেত্রে

$$\hat{Y} = \hat{\alpha} + \hat{\beta}x = \bar{y} + r\,\frac{s_y}{s_x}(x - \bar{x}) = \bar{y} \quad (\text{কারণ } r = 0)$$

অর্থাৎ নির্ভরণরেখা সাহায্যে $Y$-এর অনুমানে $x$ আমাদের কোন কাজেই আসছে না। অর্থাৎ যদি $X$-এর ওপর $Y$-এর নির্ভরণ ঋজুরৈখিক ব'লে ধরা হয়, তাহলে $X$-এর মান $Y$-এর মান নির্ণয়ে কোন আলোকপাত করতে অসমর্থ। ঠিক এ ব্যাপারটি ঘটবে $Y$-এর ওপর যদি $X$-এর নির্ভরণরেখা নির্ণয়ের চেষ্টা করা হয়, তাহলেও।

ওপরের আলোচনা থেকে এটাও স্পষ্ট যে, $r$-এর চিহ্ননিরপেক্ষ পরিমাণ অর্থাৎ $|r|$-এর পরিমাণকে $Y$-এর অনুমিতিতে (prediction) নির্ভরণরেখার

উপযোগিতার অনুমাপক হিসেবে গ্রহণ করা যায়। $|r|$-এর মান যত বেশী হবে অনুমান কার্যে নির্ভরণরেখাটির দক্ষতা ততই বেশী হবে।

(5) $$\operatorname{cov}(X, e) = \frac{1}{n}\sum(x_i - \bar{x})(e_i - \bar{e}) = \frac{1}{n}\sum(x_i - \bar{x})e_i$$

$$= \frac{1}{n}\sum x_i e_i - \bar{x}\,\frac{1}{n}\sum e_i = \frac{1}{n}\sum x_i e_i \text{ [ যেহেতু } \sum e_i = 0].$$

$$= \frac{1}{n}\sum x_i(y_i - \hat{Y}_i) = \frac{1}{n}\sum x_i(y_i - \hat{a} - \hat{\beta}x_i)$$

$= 0$ [ নর্ম্যাল সমীকরণ থেকে ]

অর্থাৎ $r_{Xe} = 0$ অর্থাৎ $e$ হচ্ছে $Y$-এর সেই অংশটুকু যা $X$-এর সঙ্গে সহগতিমুক্ত।

(6) $$\operatorname{cov}(\hat{Y}, e) = \frac{1}{n}\sum(\hat{Y} - \bar{\hat{Y}})(e_i - \bar{e})$$

$$= \frac{1}{n}\sum\left[\bar{y} + r\frac{s_y}{s_x}(x_i - \bar{x}) - \bar{y}\right]e_i$$

$$= r\frac{s_y}{s_x}\frac{1}{n}\sum(x_i - \bar{x})e_i = 0, \text{ কারণ } \operatorname{cov}(X, e) = 0.$$

সুতরাং $\operatorname{cov}(Y, \hat{Y}) = \operatorname{cov}[\hat{Y} + e, \hat{Y}]$

$$= \frac{1}{n}\sum[\{(\hat{Y} + e) - (\bar{\hat{Y}} + \bar{e})\}(\hat{Y} - \bar{\hat{Y}})]$$

$$= \frac{1}{n}\sum(\hat{Y} - \bar{\hat{Y}})^2 + \frac{1}{n}\sum(e - \bar{e})(\hat{Y} - \bar{\hat{Y}})$$

$$= V(\hat{Y}) + \operatorname{cov}(\hat{Y}, e) = V(\hat{Y}).$$

তাই $$r_{Y,\hat{Y}} = \frac{\operatorname{cov}(Y, \hat{Y})}{\sqrt{V(Y)V(\hat{Y})}} = \sqrt{\frac{V(\hat{Y})}{V(Y)}} = |r|.$$

ফলে, $Y$ এবং নির্ভরণরেখা সাহায্যে তার অনুমিত মানের সহগতি সর্বদাই অ-ঋণাত্মক।

(7) আমরা দেখেছি $b_{yx} = r\frac{s_y}{s_x}$, $b_{xy} = r\frac{s_x}{s_y}$.

সুতরাং $b_{yx} \times b_{xy} = r^2$ এবং $r = \pm\sqrt{b_{yx} \times b_{xy}}$.

কাজেই $r$ হচ্ছে $b_{yx}$ ও $b_{xy}$-এর জ্যামিতিক গড়। আবার, স্পষ্টতঃই $b_{yx}$ ও $b_{xy}$ সমচিহ্নবিশিষ্ট হবে এবং এরা যদি উভয়েই ধনাত্মক ( ঋণাত্মক ) হয়, তবে $r$ ও ধনাত্মক ( ঋণাত্মক ) হবে।

## 10.8

এখন, দুটি সম্ভাবনা চল $X$ ও $Y$-এর তত্ত্বগত বিভাজনের ভিত্তিতে একটির ওপর অপরটির নির্ভরণ সম্পর্কে সামান্য আলোকপাত করা যাক।

মনে কর $\eta_x = E(Y|x) = \psi(x)$ হচ্ছে $X$-এর ওপর $Y$-এর নির্ভরণরেখার সমীকরণ এবং একটি বিশেষ ক্ষেত্রে যথার্থই $\psi(x) = A + Bx$ ধরনের। তাহলে, $A$ ও $B$-কে $X$ ও $Y$-এর পরিঘাতের আকারে প্রকাশ করা যায়। মনে কর, $f(x, y)$, $g(x)$ ও $h(y)$ যথাক্রমে $X$ ও $Y$-এর যুগ্মবিভাজন, $X$-এর প্রান্তীয় বিভাজন ও $Y$-এর প্রান্তীয় বিভাজনের সম্ভাবনা-ঘনত্ব-অপেক্ষক, $\mu_X$, $\mu_Y$, $\sigma_X^2$, $\sigma_Y^2$ হচ্ছে তাদের গড় ও প্রমাণ-বিচ্যুতিদ্বয় ও $P$ তাদের সহগাঙ্ক। এখন যদি ধরা যায় যে, $\alpha \leqslant X \leqslant \beta$ এবং $\gamma \leqslant Y \leqslant \delta$, তবে পাওয়া যায়,

$$E(Y|x) = \int_\gamma^\delta y\left\{\frac{f(x, y)}{g(x)}\right\} dy = A + Bx.$$

ফলে, $$E(Y) = \mu_y = \int_\gamma^\delta y\, h(y)\, dy = \int_\gamma^\delta y\left[\int_\alpha^\beta f(x, y)\, dx\right] dy$$

$$= \int_\alpha^\beta \int_\gamma^\delta y\, f(x, y) dx\, dy = \int_\alpha^\beta \left(\int_\gamma^\delta y \frac{f(x, y)}{g(x)} dy\right) g(x)\, dx$$

$$= \int_\alpha^\beta E(Y|x)\, g(x)\, dx = \int_\alpha^\beta (A + Bx)\, g(x) dx$$

$$= A\int_\alpha^\beta g(x)\, dx + B\int_\alpha^\beta x\, g(x)\, dx = A + BE(X)$$

$$= A + B\mu_x.$$

আবার, $$E(XY) = \int_\alpha^\beta \int_\gamma^\delta xy\, f(x, y)\, dx\, dy$$

$$= \int_\alpha^\beta \int_\gamma^\delta xy\left(\frac{f(x, y)}{g(x)}\right) g(x)\, dx\, dy$$

$$= \int_\alpha^\beta x\left(\int_\gamma^\delta y\, \frac{f(x, y)}{g(x)} dy\right) g(x)\, dx$$

$$= \int_\alpha^\beta x\, E(Y|x)\, g(x)\, dx$$

$$= \int_\alpha^\beta x(A+Bx)\,g(x)\,dx$$

$$= A\int_\alpha^\beta x\,g(x)\,dx + B\int_\alpha^\beta x^2\,g(x)\,dx$$

$$= AE(X)+BE(X^2) = A\,\mu_X + BE(X^2).$$

সুতরাং $\text{cov}(X, Y) = E(XY) - E(X)\,E(Y)$

$$= AE(X)+BE(X^2)-AE(X)-BE^2(X)$$

$$= B\,[E(X^2)-E^2(X)] = BV(X) = B\sigma_X{}^2.$$

ফলে, $B = \dfrac{\text{cov}(X, Y)}{\sigma_x{}^2} = \dfrac{\rho\,\sigma_X\sigma_Y}{\sigma_X{}^2} = \rho\dfrac{\sigma_Y}{\sigma_X}.$

সুতরাং $A = E(Y) - B\,E(X) = E(Y) - \rho\dfrac{\sigma_Y}{\sigma_X}E(X)$

এবং $E(Y|X) = E(Y) + \rho\dfrac{\sigma_Y}{\sigma_X}(X - E(X))$

$$= \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X).$$

কাজেই, এক্ষেত্রে নির্ভরণরেখাটির সমীকরণ হচ্ছে

$$\eta_X = E(Y|X) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(X-\mu_x)$$

এবং নির্ভরণাঙ্ক হচ্ছে $\rho\dfrac{\sigma_Y}{\sigma_X}.$

পক্ষান্তরে, যদি $E(Y|x) = \psi(x)$ যথার্থই $A+Bx$ আকারের না হয়, তাহলেও যদি $\psi(x)$-এর প্রকৃত রূপ জানা না থাকে তাহলে $X$-এর প্রদত্ত মানের জন্যে $Y$-এর মান অনুমান করতে গিয়ে $\psi(x)$-কে $A+Bx$ অপেক্ষক দিয়ে পরিবর্তিত ক'রে $A+Bx$-এর মানকে $\psi(x)$-এর মানের প্রাক্কলক হিসেবে ধরার চেষ্টা করা যেতে পারে। সেক্ষেত্রে অজ্ঞাত $A$ ও $B$-এর প্রাক্কলক নির্ণয় করতে পূর্বালোচিত লঘিষ্ঠ বর্গনীতি অনুসরণ করা হয়; অর্থাৎ $X$-এর জন্যে $Y$-এর মান অনুমানে $Y = A+Bx+e = Y_x + e$ লেখা হয় যাতে $Y_x = A + B_x$ হচ্ছে $X$-এর $x$ মানের জন্যে $Y$-এর অনুমিত মান এবং $e$ হচ্ছে অবশিষ্টাংশ। এই $A$ ও $B$ এমনভাবে নির্ধারিত যে,

$S = E(e^2) = \iint e^2 f(x, y)\,dx\,dy = \iint [y - A - Bx]^2 f(x, y)\,dx\,dy$-এর

মান লঘিষ্ঠ। তাহলে, $A$ ও $B$-এর নির্ধারণে নর্ম্যাল সমীকরণ $\frac{\partial S}{\partial A}=0$ ও $\frac{\partial S}{\partial B}=0$-এর সমাধান নির্ণয় করতে হয়। এ দুটি দাঁড়ায় যথাক্রমে

$$\iint y f(x, y)\, dx\, dy = A\iint f(x, y) dx\, dy + B\iint x f(x, y)\, dx\, dy$$

অর্থাৎ $E(Y)=A+BE(X)$ ... ... (10.22)

এবং $\iint xy f(x, y)\, dx\, dy = A\iint x f(x, y)\, dx\, xy$

$$+B\iint x^2 f(x, y)\, dx\, dy$$

অর্থাৎ $E(XY)=A\,E(X)+BE\,(X^2)$. ... (10.23)

সমীকরণ (10.22) ও (10.23)-এর সমাধান ক'রে $A$ ও $B$-এর প্রাক্কলক হিসেবে পাওয়া যায়

$$\widehat{B}=\frac{E(XY)-E(X)\,E(Y)}{E(X^2)-E^2(X)}=\frac{\text{cov}\,(X, Y)}{V(X)}=\frac{\rho\,\sigma_Y\sigma_X}{\sigma_X^2}=\rho\,\frac{\sigma_Y}{\sigma_X}$$

এবং $\widehat{A}=E(Y)-\widehat{B}\,E(X)=E(Y)-\rho\frac{\sigma_Y}{\sigma_X}E(X)$.

তাহলে লঘিষ্ঠ বর্গনীতি অনুযায়ী নির্ণীত ঋজুরৈখিক নির্ভরণ সমীকরণটির রূপ দাঁড়ায়

$$Y_x=E(Y)+\rho\,\frac{\sigma_Y}{\sigma_X}(x-E(X)).$$

ঠিক একইভাবে দেখানো যায় যে, $Y$-এর ওপর $X$-এর নির্ভরণ অপেক্ষক $\xi_y=E(X|y)=\phi(y)$ যদি আসলে ঋজুরৈখিক হয়, অর্থাৎ যদি $\phi(y)=c+Dy$ হয়, তবে

$\xi_y=E(X)+\rho\,\frac{\sigma_X}{\sigma_Y}(y-E(Y))$ হবে এবং যদি আসল নির্ভরণ অপেক্ষক ঋজুরৈখিক না হয় তাহলে লঘিষ্ঠ বর্গনীতি অনুযায়ী অনুমিত ঋজুরৈখিক নির্ভরণ অপেক্ষকেরও গঠন ঠিক এরকমই হবে।

## 10.9 দ্বিচল নর্ম্যাল বিভাজন (Bivariate Normal Distribution) :

আমরা এখন একটি প্রয়োজনীয় দ্বিচল তত্ত্বগত সম্ভাবনা বিভাজনের উদাহরণ দেবো। এটিকে বলে দ্বিচল নর্ম্যাল বিভাজন।

মনে কর, $X$ ও $Y$ দুটি অবিচ্ছিন্ন সম্ভাবনা চল এবং তাদের যুগ্মবিভাজনের সম্ভাবনা-ঘনত্ব-অপেক্ষক $f(x, y)$ হচ্ছে নিম্নরূপ :

$$f(x, y) = k\exp\left[-(Ax^2 + Bxy + Cy^2)\right], \qquad -\infty < x < +\infty, \; -\infty < y < +\infty.$$

এখানে, $A$, $B$ ও $C$ হচ্ছে তিনটি ধ্রুবক যাদের প্রকৃতি এমন যে, $(x, y)$-এর $(0, 0)$ ব্যতীত সব মানের জন্যেই $Ax^2 + Bxy + Cy^2 > 0$. এখানে এই দ্বিচলবিশিষ্ট প্রকাশন $Ax^2 + Bxy + Cy^2$-কে দ্বিঘাতরূপ (Quadratic Form) বলা হয় এবং

$$x \neq 0, \; y \neq 0 \text{ হলে } Ax^2 + Bxy + Cy^2 > 0 \qquad \cdots \quad (10.24)$$

এই সর্তাধীন দ্বিঘাতরূপটিকে বলা হয় ধনাত্মক দ্বিঘাতরূপ (Positive Definite Quadratic Form). এই সর্তের প্রয়োজন হচ্ছে এই যে, $Ax^2 + Bxy + Cy^2 < 0$ হলে $\int f(x, y)\, dx\, dy$-এর মান সসীম হবে না ও ফলে $f(x, y)$ কোন সম্ভাবনা-ঘনত্ব-অপেক্ষক হতে পারবে না। এছাড়া $Ax^2 + Bxy + Cy^2 = 0$ হলে $f(x, y)$-এর মান সব $x$ ও $y$-এর জন্যেই ধ্রুবক $(=k)$ হবে এবং অনাবশ্যক বোধে সেই পরিস্থিতিটি আমরা আলোচনা থেকে বাদ দেব।

$f(x, y)$ অপেক্ষকে উল্লিখিত ধ্রুবক $k$-এর মান এরূপে স্থিরীকৃত যেন

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1 \text{ হয়।} \qquad \cdots \quad (10.25)$$

এইভাবে $f(x, y)$-এর সাহায্যে ওপরে যে সম্ভাবনা বিভাজনটি নির্দিষ্ট হ'ল তাকে বলা হয় দ্বিচল নর্ম্যাল বিভাজন। এখন $X$ ও $Y$-এর প্রান্তীয় বিভাজনের ঘনত্ব-অপেক্ষক, গড়, প্রমাণ-বিচ্যুতি ও সহগাঙ্ক যথাক্রমে $g(x)$, $h(y)$, $\mu_X$, $\mu_Y$, $\sigma_X^2$, $\sigma_Y^2$ ও $\rho$ দ্বারা চিহ্নিত ক'রে প্রমাণ করা যায় যে, $K$, $A$, $B$ ও $C$-এর মান এমনভাবে এদের আকারে প্রকাশ করা যাবে যাতে $f(x, y)$-কে লেখা যাবে

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \cdot \exp.\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right\}\right], \qquad -\infty < x < \infty, \; -\infty < y < \infty. \qquad \cdots \quad (10.26)$$

## 10.10 দ্বিচল নর্ম্যাল বিভাজনের কয়েকটি ধর্ম
(Some properties of the bivariate normal distribution) :

1. দ্বিচল নর্ম্যাল বিভাজনের সম্ভাবনা-ঘনত্ব-অপেক্ষক $f(x, y)$-এর রূপ (10.26)-এর আকারে প্রকাশযোগ্য।

2. $X$ ও $Y$-এর যুগ্ম সম্ভাবনা বিভাজন নর্ম্যাল হলে $X(Y)$-এর প্রান্তীয় বিভাজন একচল নর্ম্যাল (univariate normal) হবে। আরও বিশদভাবে দেখানো যাবে যে, $X$ ও $Y$-এর সম্ভাবনা-ঘনত্ব-অপেক্ষক যথাক্রমে

$$g(x) = \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sigma_X} \exp\left[-\frac{1}{2\sigma_X^2}(x-\mu_X)^2\right], \qquad -\infty < x < \infty$$

এবং $$h(y) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_Y} \exp.\left[-\frac{1}{2\sigma_Y^2}(y-\mu_Y)^2\right], \qquad -\infty < y < \infty.$$

3. $X(Y)$-এর মান কোন নির্দিষ্ট অন্তরে রয়েছে এমন সর্তাধীনে $Y(X)$-এর বিভাজন নর্ম্যাল হবে। এই নিবেশন দুটির সম্ভাবনা-ঘনত্ব-অপেক্ষকদ্বয়ের আকার হবে যথাক্রমে নিম্নরূপ। $Y$-এর সর্তাধীন বিভাজনের সম্ভাবনা-ঘনত্ব-অপেক্ষকের $x$ বিন্দুতে গৃহীত মান হচ্ছে

$$f(y/x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_Y\sqrt{1-\rho^2}} \exp.\left[-\frac{1}{2(1-\rho^2)\sigma_Y^2}\left\{y-\mu_Y-\rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X)\right\}^2\right] \qquad -\infty < y < \infty$$

এবং $X$-এর সর্তাধীন বিভাজনের সম্ভাবনা-ঘনত্ব-অপেক্ষকের $y$ বিন্দুতে গৃহীত মান হচ্ছে

$$f(x/y) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_X} \frac{1}{\sqrt{1-\rho^2}} \exp.\left[-\frac{1}{2(1-\rho^2)\sigma_X^2}\left\{x-\mu_X-\rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)\right\}^2\right], \qquad -\infty < x < \infty.$$

4. দেখানো যায় যে, $X(Y)$-এর ওপর $Y(X)$-এর নির্ভরণ অপেক্ষক ঋজুরৈখিক।

আরও স্পষ্টভাবে,

$$\eta_x = E(Y/x) = \int_{-\infty}^{\infty} yf(y/x)dy = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X),$$

এবং $$\xi_y = E(X/y) = \int_{-\infty}^{\infty} xf(x/y)dx = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y).$$

এখানে $E(Y/x)$ $[E(X/y)]$, $X(Y)$ চলের $x(y)$ বিন্দুতে $Y(X)$-এর সর্তাধীন গাণিতিক প্রত্যাশা।

5. দেখানো যায় যে, $\sigma_{Y/x}{}^2 = \int_{-\infty}^{\infty} (y - E(Y/x))^2 f(y/x)dx$

$$= \sigma_Y{}^2(1 - \rho^2)$$

এবং $\sigma_{X/y}{}^2 = \int_{-\infty}^{\infty} \{x - E(X/y)\}^2 f(x/y)dx = \sigma_X{}^2(1 - \rho^2)$.

এখানে $\sigma^2{}_{Y/x}$ $(\sigma_{X/y}{}^2)$-কে বলা হয় $X(Y)$ চলের মান $x(y)$ বিন্দুতে নিবদ্ধ থাকার সর্তাধীন ভেদমান। উল্লেখ করা যেতে পারে যে, $\sigma^2{}_{Y/x}$ $(\sigma^2{}_{X/y})$-এর মান সব $x(y)$-এর জন্যেই ধ্রুবক। এই ধর্মকে বলে প্রভেদ ধ্রুবকত্ব (homoscedasticity) এবং এই ধর্মের অস্তিত্ব থাকার জন্যে উল্লিখিত সর্তাধীন বিভাজনকে প্রভেদ ধ্রুবকত্বসম্পন্ন (homoscedastic) ব'লে অভিহিত করা হয়।

6. দ্বিচল নর্ম্যাল বিভাজনের ক্ষেত্রে $\rho$ একটি সার্থক সহগাঙ্ক, কারণ যদি $\rho = 0$ হয়, তবে প্রত্যেক $x$ ও $y$-এর জন্যেই

$$f(x, y) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp.\left[-\frac{1}{2\sigma_X{}^2}(x - \mu_X)^2\right]$$

$$\frac{1}{\sqrt{2\pi}\sigma_Y} \exp.\left[\frac{1}{2\sigma_Y{}^2}(y - \mu_Y)^2\right]$$

$= g(x)\, h(y)$ অর্থাৎ সম্ভাবনা তত্ত্বানুযায়ী $X$ ও $Y$ পরস্পর অনধীন। আবার, যদি $\rho = \pm 1$ হয়, তাহলে $\sigma_{Y/x}{}^2 = \sigma_Y{}^2(1 - \rho^2) = 0$ অর্থাৎ $E[Y - E(Y/x)]^2 = 0$ অর্থাৎ প্রত্যেক $x$-এর জন্যেই $P[Y = E(Y/x)] = 1$. ফলে, $X$-এর প্রত্যেক $x$ মানের জন্যে নির্দিষ্ট $Y$ স্তবকের মানই সেই স্তবকের গড়ের সমান হওয়ার সম্ভাবনা হচ্ছে 1. তার ফল হচ্ছে এই যে, যেহেতু

$$E(Y/x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

এবং প্রত্যেক $x$-এর জন্যে $P[Y = E(Y/x)] = 1$, কাজেই $Y$ চলটি $X$ চলের একটি ঋজুরৈখিক অপেক্ষক হওয়ার সম্ভাবনা হচ্ছে 1.

আবার $\rho$-এর মান $\pm 1$-এর কাছাকাছি হওয়ার সঙ্গে সঙ্গে $\sigma^2{}_{Y/x}$-এর মান 0-এর কাছাকাছি হতে থাকে। তাই বলা যায় যে, $\rho$ যতই $\pm 1$-এর কাছাকাছি যাবে $X(Y)$-এর যে কোন নির্দিষ্ট মান $x(y)$-এর জন্যে $Y(X)$-এর সর্তাধীন সম্ভাবনা বিভাজনের বিস্তৃতি ততই হ্রাস পাবে।

**10.10 (a) সংস্রব মাপনায় সহগাঙ্কের ব্যর্থতা (Failure correlation coefficient in measuring association) :**

আমরা আগে একটি উদাহরণে দেখেছি যে, সহগাঙ্কের মান শূন্য হলেও দুটি চল পরস্পর নির্ভরশীল এমন কি কোন গাণিতিক সম্পর্কসূত্রেও আবদ্ধ হতে পারে। তা থেকেই বোঝা যায় যে, সহগাঙ্ক $r$ চলদুটির সব রকম সংস্রবের রূপ বা প্রকৃতি প্রকাশে সমর্থ নয়। বাস্তবিক, যদি চলদুটির সম্পর্ক নিশ্চিতভাবে অ-ঋজুরৈখিক হয়, তবে সে জাতীয় সংস্রবের অস্তিত্ব সহগাঙ্কের মাধ্যমে ধরা পড়ে না। উদাহরণস্বরূপ বলা যায় যে, $Y$ ও $X$ চলদুটির মধ্যে যদি একটি দ্বিঘাত-ভিত্তিক গাণিতিক সম্পর্ক থাকে [ উদাহরণতঃ $Y$ যদি $X$-এর দ্বিঘাতজ অপেক্ষক হয় ], তাহলে $r$-এর মান অনেক সময় শূন্য বা নগণ্য পরিমাণ হয়ে থাকে, যার ফলে $r$-এর সাহায্যে তাদের প্রকৃত সংস্রব সম্পর্কে ধারণা করা যায় না। পক্ষান্তরে, চলদুটির সংস্রব যদি এমন হয় যে, তাদের পারস্পরিক সম্পর্ককে মোটামুটিভাবে একটি ঋজুরৈখিক কাঠামোর মধ্যে ফেলা যায়, তাহলে তাদের ঐ জাতীয় অন্ততঃ আসন্নভাবে ঋজুরৈখিক সম্পর্কসূত্র সহগাঙ্কের সাহায্যে সার্থকভাবে পরিমাপ করার চেষ্টা করা যায়। এ মন্তব্যের যাথার্থ্য সহগতি ও ঋজুরৈখিক নির্ভরণ তত্ত্বের আলোচনা থেকে কিছুটা উপলব্ধি করা যাবে। আমরা দেখেছি যে, নির্ভরণ যদি অন্ততঃ আসন্নভাবেও ঋজুরৈখিক প্রকৃতির হয়, তবে সহগাঙ্কের মানের সাহায্যে ঐ নির্ভরণের গুরুত্ব বা লঘুত্ব বিচার করা যায়। সহগাঙ্কের মান খুব কম হলে এটাই প্রমাণ হবে যে, একটি চলের অপরটির ওপর নির্ভরণ ঋজুরেখার সাহায্যে প্রকাশ সার্থকভাবে করা যায় না এবং $r$-এর চিহ্নিনিরপেক্ষ মান যদি সর্বাধিক অর্থাৎ যদি $|r| = 1$ হয়, তাহলে একটি চলের ওপর অপরটির নির্ভরণ বাস্তবিক ঋজুরৈখিক হবে ও $r$-এর অন্তর্বর্তী পরিমাপগুলির জন্যে $|r|$-এর মান 1 থেকে যতদূরে হবে ঐ নির্ভরণের প্রকৃতি ঋজুরৈখিক প্রকৃতি থেকে ততদূরে হবে। কাজেই $|r|$-এর মান যদি খুব অল্প হয়, তাহলে সার্থকভাবে এটাই বলা যাবে যে, চলদুটির মধ্যে ঋজুরৈখিক সংস্রব থাকার সম্ভাবনা খুব কম। তাদের মধ্যে কোন অ-ঋজুরৈখিক ঘনিষ্ঠ সম্পর্ক থাকতে পারে অথবা তাদের মধ্যে কোন সংস্রব না থাকতেও পারে। এসব ব্যাপার সম্পর্কে উপযুক্ত সিদ্ধান্তে উপনীত হতে হলে বিক্ষেপণ চিত্রটি ভালোভাবে পর্যবেক্ষণ করা দরকার। চিত্র থেকে যদি চলদুটির মধ্যে অন্ততঃ আংশিকভাবেও ঋজুরৈখিক সম্পর্কের সূত্র ধরা না পড়ে, তাহলে তাদের সংস্রব খুঁজতে $r$-এর

ওপর নির্ভর না করাই উচিত। আবার $r$-এর মান কম হলেও যদি ঐ চিত্র থেকে চলদুটির মধ্যে কোন সম্পর্কসূত্র আঁচ করা যায়, যেমন উদাহরণ-স্বরূপ, $Y=A+BX+CX^2$ বা ঐ রকম অন্য কোন অ-ঋজুরৈখিক সম্পর্ক অনুমান করা যায়, তাহলে $|r|$-এর মান ছোট হওয়া সত্ত্বেও মোটামুটি বলা যেতে পারে যে, চলদুটির মধ্যে সংশ্রব থাকা সম্ভব এবং এ জাতীয় অ-ঋজুরৈখিক সংশ্রব মাপনের অন্য পদ্ধতি অবলম্বন করার চেষ্টা করা উচিত।

একটি কথা এসম্পর্কে সর্বদাই মনে রাখা উচিত যে, সহগতির আলোচনাসূত্রে চলদুটির মধ্যে কার্যকারণ সূত্র (cause-effect relationship) আবিষ্কারের চেষ্টা করা ভুল। সহগতি খুব বেশী হলে এমন সিদ্ধান্ত করা যাবে যে, চলদুটির মধ্যে খুবই ঘনিষ্ঠ সংশ্রব রয়েছে। কিন্তু একটি চলের পরিবর্তনই অন্যটির পরিবর্তনের জন্যে দায়ী অর্থাৎ একটি চল অপর চলটির বিভিন্ন মান গ্রহণের কারণ বা হেতু এরকম সিদ্ধান্ত করা অনেকসময়ই ভ্রমাত্মক। বাস্তবিক, এরকম ঘনিষ্ঠ সংশ্রবযুক্ত দুটি চলের মান গ্রহণই অপর অনালোচিত তৃতীয় কোন চল বা একাধিক অজ্ঞাত চলের প্রভাবে ঘটতে পারে। সহগতির আলোচনায় এজাতীয় প্রয়োগসীমার কথা সর্বদা মনে রাখা দরকার। যেমন, কয়েকজন গৃহস্বামীর বাড়ীভাড়া বাবদ এবং বিলাসদ্রব্যের ওপর ব্যয়ের হিসেব নিলে দেখা যাবে যে, এর একটির বাড়লে অপরটিরও বাড়ছে। ফলে, মনে করা যেতে পারে যে, এই দুটি খাতের ব্যয় পরস্পর ধনাত্মক সহগতিযুক্ত। কিন্তু এমন হওয়া স্বাভাবিক যে গৃহস্বামীদের মোট আয়বৃদ্ধির জন্যেই ঐ দুটি খাতে ব্যয়ের পরিমাণও বাড়ছে।

## 10.11 সহগতি অনুপাত (Correlation Ratio) :

দুটি পরস্পর সংশ্রবযুক্ত চল $X$ ও $Y$-এর একটি (ধর $Y$) যদি অপরটির (অর্থাৎ $X$-এর) ওপর নির্ভরশীল হয়, তাহলে আমরা আগে দেখেছি যে, নির্ভরণ নীতি প্রয়োগ ক'রে দ্বিতীয়টির মান থেকে প্রথমটি সম্পর্কে অনুমান করা যায় এবং ঐ নির্ভরণ যদি অন্ততঃ আসন্নভাবেও ঋজুরৈখিক হয় তাহলে ঐ নির্ভরশীলতার পরিমাণ মাপা যায় $|r|=\sqrt{\frac{V(Y_X)}{V(Y)}}$ এই মাপনাঙ্কটির সাহায্যে।

এখানে $Y_x=a+bx=\bar{y}+r\frac{s_y}{s_x}(x-\bar{x})$ হচ্ছে লঘিষ্ঠ বর্গনীতি অনুযায়ী নির্ধারিত $X$-এর ওপর $Y$-এর সাযুজ্য রক্ষাকারী নির্ভরণ ঋজুরেখার সমীকরণ। এখন,

$X$-এর ওপর $Y$-এর নির্ভরতা মাপনকার্যে $|r|$-এর উপযোগিতা তখনই স্বীকার্য হবে যখন এই নির্ভরতা অন্ততঃ আসন্নভাবেও ঋজুরৈখিক ব'লে ধরা যায়। কিন্তু যদি ঐ নির্ভরতা আদৌ ঋজুরৈখিক প্রকৃতির না হয় তখন অবশ্যই অন্য কোন মাপনাঙ্কের অনুসন্ধান করা প্রয়োজন। বাস্তবিক, সেক্ষেত্রে ঐরকম একটি মাপনাঙ্ক সম্পর্কে এখন আমরা আলোচনা ক'রব।

ধর, $X$-এর বিভিন্ন মানগুলি হচ্ছে $x_1, \ldots x_i, \ldots, x_k$ এবং $Y$-এর যে সমস্ত মানের জন্যে ব্যষ্টিগুলির $X$ মান হচ্ছে $x_i$, সেগুলি মনে কর $y_{i1}, \ldots, y_{in_i}$ $(i=1, \ldots, k)$; তাহলে প্রদত্ত রাশিমালা হচ্ছে $x_i$ $(i=1, \ldots, k)$ এবং $y_{ij}$ $(j=1, \ldots, n_i\,;\ i=1, \cdots, k)$ যার মোট সংখ্যা হচ্ছে $n=\sum_{i=1}^{k} n_i$.

এখন, $X=x_i$-এর জন্যে $y_{ij}$ $(j=1, \ldots, n_i)$—

এই $n_i$ সংখ্যক মানগুলি মনে কর, একটি স্তবক বা পঙক্তি গঠন করেছে। তাহলে আমরা মোট $k$ সংখ্যক স্তবক বা পঙক্তি পেলাম। এখন, $i$-তম পঙক্তিস্থিত $Y$ মানগুলির গড় হচ্ছে $\bar{y}_i=\frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$. এখন, $x_i$-এর পরিবর্তনের সঙ্গে সঙ্গে $\bar{y}_i$ মানগুলি কিভাবে পরিবর্তিত হয় $X$-এর ওপর $Y$-এর নির্ভরণ সেটিকেই প্রকাশ করে। কাজেই $X$-এর ওপর $Y$-এর নির্ভরতা মাপনে $Y_x$ ও $V(Y_x)$-কে বিবেচনা করার পরিবর্তে $\bar{y}_i$ এবং তার ভেদমান $V(\bar{y}_i)$-এর মাধ্যমে $r^2=\frac{V(Y_x)}{V(Y)}$-এর অনুরূপ $\frac{V(\bar{y}_i)}{V(Y)}$ এই অনুপাতটিকে গ্রহণ করা উচিত।

এখন, $\bar{y}_i$ মানগুলির গড় হচ্ছে

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{k} n_i\,\bar{y}_i=\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij},$$

$$V(\bar{y}_i)=\frac{1}{n}\sum_{i=1}^{k} n_i\,(\bar{y}_i-\bar{y})^2,$$

এবং $$V(Y)=\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2=s_y^{\,2}.$$

কাজেই $$\eta_{yx}=\sqrt{\frac{V(\bar{y}_i)}{V(Y)}}=\frac{\sqrt{\sum n_i(\bar{y}_i-\bar{y})}}{\sqrt{\sum\sum(y_{ij}-\bar{y})^2}}$$

এই মাপনাঙ্কটিই হচ্ছে $X$-এর ওপর $Y$-এর সহগতি অনুপাতের সংজ্ঞা এবং এর সাহায্যেই $X$-এর ওপর $Y$-এর নির্ভরশীলতা মাপা হয়ে থাকে, বিশেষতঃ যখন নিশ্চিতভাবে জানা যায় যে, এই নির্ভরতা অন্ততঃ মোটামুটিভাবেও ঋজুরৈখিক প্রকৃতির নয়।

উল্লেখযোগ্য যে $X$ ও $Y$ যদি সম্ভাবনা চল হয় এবং তাদের ঔপপত্তিক বিভাজনের স্বরূপ যদি জানা থাকে তাহলে $\sqrt{\frac{V(E(Y|X))}{V(Y)}}$-কে $X$-এর ওপর $Y$-এর সহগতি অনুপাতের সংজ্ঞা হিসেবে ধরা যায়।

এখানে $V(y)=E[Y-E(y)]^2$

এবং $V[E(Y|X)]=E[E(Y|X)-E(Y)]^2$,

কারণ $E[E(Y|X)]=E(Y)$. এখানে $E(Y|X)$-কে $Y$-এর সর্তাধীন গাণিতিক প্রত্যাশা বলা হয় এবং এটি নিজেই একটি সম্ভাবনা চল।

**10.12 সহগতি অনুপাতের কয়েকটি ধর্ম** (Some properties of correletion ratio) :

$$\sum_i\sum_j(y_{ij}-\bar{y})^2=\sum\sum[(y_{ij}-\bar{y}_i)+(\bar{y}_i-\bar{y})]^2$$

$$=\sum\sum(y_{ij}-\bar{y}_i)^2+\sum n_i(\bar{y}_i-\bar{y})^2$$

$$+\sum_i(\bar{y}_i-\bar{y})\sum_j(y_{ij}-\bar{y}_i)$$

সুতরাং $$\frac{1}{n}\sum_i\sum_j(y_{ij}-\bar{y}_i)^2=\frac{1}{n}\sum_i\sum_j(y_{ij}-\bar{y})^2$$

$$-\frac{1}{n}\sum n_i(\bar{y}_i-\bar{y})^2$$

$\left[\text{যেহেতু } \sum(y_{ij}-\bar{y}_i)=0\right.$

$$= s_y^2 - s_y^2 \eta^2_{yx} = (1-\eta^2_{yx})\, s_y^2$$

অর্থাৎ $s_y^2 = \eta^2_{yx}\, s_y^2 + (1-\eta^2_{yx})\, s_y^2.$

আবার, $Y_i = a + b\,x_i$ লিখলে এবং $a$ ও $b$ যদি

$\sum n_i\,(\bar{y}_i - Y_i) = 0$ ও $\sum n_i x_i\,(\bar{y}_i - Y_i) = 0$ এই দুটি নর্ম্যাল সমীকরণের সমাধানযোগে নির্ণীত হয়, তাহলে পাওয়া যায়

$$Y_i = \bar{y} + r\,\frac{s_y}{s_x}\,(x_i - \bar{x}).$$

ফলে, $$\sum n_i\,(\bar{y}_i - \bar{y})^2 = \sum n_i[(\bar{y}_i - Y_i) + (Y_i - \bar{y})]^2$$

$$= \sum n_i\,(\bar{y}_i - Y_i)^2 + \sum n_i\,(Y_i - \bar{y})^2$$

$$+ 2\sum n_i(\bar{y}_i - Y_i)(Y_i - \bar{y}).$$

কিন্তু $$\sum n_i(\bar{y}_i - Y_i)(Y_i - \bar{y}) = \sum n_i(\bar{y}_i - Y_i)(a + bx_i - \bar{y})$$

$$= (a - \bar{y}) \sum n_i(\bar{y}_i - Y_i) + b \sum n_i x_i(\bar{y}_i - Y_i)$$

$= 0$ [ নর্ম্যাল সমীকরণ দুটি ব্যবহার ক'রে ].

আবার, $\frac{1}{n}\sum n_i\,(Y_i - \bar{y})^2 = V(Y_x) = r^2 s_y^2 =$ ঋজুরৈখিক নির্ভরণ-জনিত ভেদমান।

সুতরাং $$\frac{1}{n}\sum n_i(\bar{y}_i - \bar{y})^2$$

$$= \frac{1}{n}\sum n_i(\bar{y}_i - Y_i)^2 + \frac{1}{n}\sum n_i(Y_i - \bar{y})^2$$

থেকে পাওয়া যায়

$$\frac{1}{n}\sum n_i(y_i - Y_i)^2 = \frac{1}{n}\sum n_i\,(\bar{y}_i - \bar{y})^2 - \frac{1}{n}\sum n_i(Y_i - \bar{y})^2$$

অর্থাৎ $$\frac{1}{n}\sum n_i(\bar{y}_i - Y_i)^2 = \eta^2_{yx} s_y^2 - r^2 s_y^2 = (\eta^2_{yx} - r^2) s_y^2$$

এবং $$\eta^2_{yx} s_y^2 = r^2 s_y^2 + (\eta^2_{yx} - r^2) s_y^2.$$

তাহলে, স্পষ্টতঃই, যেহেতু $\sum\sum (y_{ij} - \bar{y}_i)^2 > 0$

এবং $\sum n_i(\bar{y}_i - Y_i)^2 \geqslant 0$, কাজেই আমরা পাই

$$1-\eta^2_{yx} \geqslant 0 \quad \text{এবং} \quad \eta^2_{yx}-r^2 \geqslant 0$$

অর্থাৎ $r^2 \leqslant \eta_{yx}^2 \leqslant 1$.

অধিকন্তু, $\eta^2_{yx}=r^2$, যদি প্রত্যেক $i=1,\ldots, k$-এর জন্যে $\bar{y}_i - Y_i=0$ অর্থাৎ $Y_i=\bar{y}_i$ হয়, অর্থাৎ যদি $X$-এর ওপর $Y$-এর নির্ভরণ রেখা ঋজু হয়। কাজেই $(\eta^2_{yx}-r^2)$-কে দেখা যেতে পারে $X$-এর ওপর $Y$-এর নির্ভরণের প্রকৃতি ঋজুরৈখিক প্রকৃতি থেকে কতটা ভিন্ন তার একটি মাপনাঙ্ক হিসেবে। সহগতি অনুপাতের এটি আর একটি উপযোগিতা ও গুরুত্ব। আবার, $r^2$-কে যেমন দেখা যায় $Y$-এর মোট প্রভেদের যতটুকু $X$-এর ওপর $Y$-এর ঋজুরৈখিক নির্ভরণ-মাধ্যমে ব্যাখ্যাত হয়েছে তেমনিভাবে $\eta^2_{yx}$ কেও দেখা যেতে পারে $Y$-এর মোট প্রভেদের যতটুকু প্রদত্ত $X$-গুলির জন্যে $Y$-এর পঙক্তিগড়গুলির প্রভেদের সূত্রে ব্যাখ্যাত হয়েছে।

ঠিক যেমন $X$-এর ওপর $Y$-এর সহগতি ভগ্নাংশ নির্দিষ্ট করা হয়েছে, তেমনি $Y$-এর ওপর $X$-এর সহগতি অনুপাত $\eta_{xy}$-কেও আমরা নির্দেশ করতে পারি।

## 10.13 মানক্রমিক সহগতি (Rank correlation):

ধর $n$-সংখ্যক ব্যষ্টি রয়েছে যাদের সম্পর্কে দুটি বিভিন্ন চরিত্রবৈশিষ্ট্য আলোচনা করতে হবে এবং মনে কর ঐ চরিত্রবৈশিষ্ট্য-দুটি সংখ্যাযোগে মাপন-যোগ্য নয়। উদাহরণস্বরূপ মনে কর, আমরা দেখতে চাই (1) সপ্রতিভতা ও (2) শিল্পরসবোধ এই দুটি গুণ বা চরিত্রবৈশিষ্ট্য কোন $n$-সংখ্যক ব্যক্তির মধ্যে কী রকম বিভিন্ন মাত্রায় রয়েছে।

এসব ক্ষেত্রে চরিত্র-দুটির মধ্যে সংস্রব আছে কিনা তা দেখবার জন্যে স্বভাবতঃই পূর্বে আলোচিত সহগাঙ্ক ব্যবহার করা অসম্ভব। এক্ষেত্রে আমরা একটি নতুন ধরনের মাপনাঙ্ক (cofficient) ব্যবহার ক'রে এদের সংস্রব মাপনের পদ্ধতি আলোচনা ক'রব। এই মাপনাঙ্ককে বলে মানক্রমিক সহগাঙ্ক (rank correlation coefficient) যার প্রকৃতি এখন বিশ্লেষণ করা হবে। এর প্রয়োগ অবশ্য ব্যাপকতর করা যায়। অনেক সময়, চরিত্রবৈশিষ্ট্য এমন হতে পারে যে, তাদের পরিমাপ করা যায়, কিন্তু তাতে অনেক সময় ও অর্থ ব্যয় করতে হয় এবং তা এড়াবার জন্যে এদের পরিমাণ নির্ণয় করা হয় না। দ্বিতীয়তঃ এদের

আসল মানগুলি জানা থাকলেও কখনও কখনও পূর্বালোচিত মাপনাঙ্ক $r$ ব্যবহার না ক'রে তার পরিবর্তে সময় সংক্ষেপের প্রয়োজনে আলোচ্য নতুন মাপনাঙ্কটি ব্যবহার করা হয়।

মনে কর $A$ ও $B$ দুটি চরিত্রবৈশিষ্ট্য $n$-সংখ্যক বিভিন্ন ব্যষ্টির মধ্যে ভিন্ন ভিন্ন মাত্রায় রয়েছে এবং সে মাত্রার ক্রম অনুযায়ী তাদেরকে পরপর সাজানো যায়; অর্থাৎ $n$ সংখ্যক ব্যষ্টির মধ্যে $A$-চরিত্রবৈশিষ্ট্যটি কার মধ্যে সবচেয়ে বেশী আছে এবং তার পরবর্তী মাত্রায় কার মধ্যে আছে ইত্যাদি এবং সবশেষে সবচেয়ে স্বল্পমাত্রায় কার মধ্যে আছে তা নির্ণয় করা যায়। সে অনুযায়ী মনে কর $A$-চরিত্রটি বিভিন্ন মাত্রায় অধিকার করার সূত্রে $n$ সংখ্যক ব্যষ্টিকে পরপর $n$ সংখ্যক অনুক্রম মান (rank) দেওয়া হ'ল; অর্থাৎ সর্বোচ্চ মাত্রাধিকারীকে 1, তৎপরবর্তী মাত্রাধিকারীকে 2, ইত্যাদি এবং সর্বনিম্ন মাত্রাধিকারীকে অনুক্রমমান $n$ দেওয়া হ'ল। তাহলে কোন ব্যষ্টিকে অনুক্রম মান $t$ আরোপ করা হবে যদি $(t-1)$ সংখ্যক ব্যষ্টির মধ্যে $A$-চরিত্রটি অধিকতর মাত্রায় বিদ্যমান হয় $(t=1, 2,..., n)$. মনে কর, $A$-চরিত্রানুযায়ী, $n$ সংখ্যক ব্যষ্টির অনুক্রম মান হ'ল $u_1,......u_i,......u_n$. এখানে প্রত্যেক $i=1,......, n$-এর জন্যে $u_i$ হচ্ছে $1, 2,..., n$-এর মধ্যবর্তী যে কোন একটি সংখ্যা এবং $i \neq j$ হলে $u_i \neq u_j$. তেমনিভাবে, মনে কর, $B$-চরিত্রবৈশিষ্ট্যানুযায়ী ঐ $n$-সংখ্যক ব্যষ্টির অনুক্রমমান যথাক্রমে $v_1,...v_i,..., v_n$. এই $v_i$ $(i=1,..., n)$ সংখ্যাগুলিও $1, 2,..., n$—এই সংখ্যাগুলিরই এক একটি সংখ্যা এবং $i \neq j$ হলে $v_i \neq v_j$. এখন, $A$ ও $B$ চরিত্র-দুটির মধ্যে সংশ্রব আছে কিনা তা জানার জন্যে $U$ ও $V$ এই দুটি মাপনযোগ্য চলের মধ্যে যে সহগতি আছে তা নির্ণয় করা যেতে পারে। এখানে $U$ ও $V$ হচ্ছে যথাক্রমে $u_i(i=1, ..., n)$ ও $v_i(i=1, ..., n)$ মান-গ্রহণকারী চলদ্বয়। এখন, $U$ ও $V$-এর সহগতিকে বলে $A$ ও $B$ চরিত্র-দুটির মানক্রমিক সহগতি (rank correlation). $U$ ও $V$-এর সহগাঙ্ক $r_{uv}$ দ্বারা সূচিত করলে $R_{AB}$ দ্বারা সূচিত করা হয় $A$ ও $B$-এর মানক্রমিক সহগাঙ্ককে এবং আমরা লিখব

$$R_{AB}=r_{uv}=\frac{\text{cov}\,(U, V)}{\sqrt{V(U)}\,\sqrt{V(V)}}.$$

এখন, যদি দেখা যায় যে, প্রত্যেক $i=1, ..., n$-এর জন্যে $u_i=v_i$ অর্থাৎ যদি $A$ ও $B$-এর সূত্রে অনুক্রম মানগুলির মধ্যে সম্পূর্ণ মিল থাকে, তবে বলা

হবে যে, $A$ ও $B$ চরিত্র-দুটি ধনাত্মকভাবে সম্পূর্ণ সংশ্রবযুক্ত। পক্ষান্তরে, যদি প্রত্যেক $i=1, \ldots, n$-এর জন্যে $v_i = n - u_i + 1$ হয়, অর্থাৎ যদি $u_i$ বাড়লে $v_i$-এর মান ক্রমাগত কমতে থাকে অর্থাৎ যদি $U$ ও $V$-এর মধ্যে পূর্ণ অমিল বা বৈপরীত্য থাকে, তাহলে বলা উচিত যে, $A$ ও $B$ হচ্ছে ঋণাত্মকভাবে সম্পূর্ণ সংশ্রবযুক্ত। $U$ ও $V$ চল-দুটির মধ্যে অন্যান্য অন্তর্বর্তী সম্পর্কের জন্যে $r_{uv}$-এর সাহায্যে $A$ ও $B$-এর সংশ্রব মাপা যেতে পারে।

মনে কর $d_i = u_i - v_i$, $i=1, \ldots, n$. তাহলে,

$$\bar{u} = \frac{1}{n}\sum_{i=1}^{n} u_i = \frac{1}{n}(1+2+\cdots+i+\cdots+n) = \frac{n+1}{2},$$

$$\bar{v} = \frac{1}{n}\sum_{i=1}^{n} v_i = \frac{1}{n}(1+2+\cdots+i+\cdots+n) = \frac{n+1}{2},$$

অর্থাৎ, $$\bar{u} = \bar{v},\ V(U) = \frac{1}{n}\sum (u_i - \bar{u})^2 = \frac{n^2-1}{12}$$

$$= V(V) = \frac{1}{n}\sum (v_i - \bar{v})^2.$$

আবার, $$\frac{1}{n}\sum_{i=1}^{n} d_i^2 = \frac{1}{n}\sum (u_i - v_i)^2$$

$$= \frac{1}{n}\sum \{(u_i - \bar{u}) - (v_i - \bar{v})\}^2$$

$$= \frac{1}{n}\sum (u_i - \bar{u})^2 + \frac{1}{n}\sum (v_i - \bar{v})^2 - \frac{2}{n}\sum (u_i - \bar{u})(v_i - \bar{v})$$

$$= V(U) + V(V) - 2\,\mathrm{cov}\,(U, V) = 2\,V(U) - 2\,\mathrm{cov}\,(U, V).$$

সুতরাং, $$\mathrm{cov}\,(U, V) = V(U) - \frac{1}{2n}\sum d_i^2.$$

তাই, $$r_{uv} = \frac{\mathrm{cov}\,(U, V)}{V(U)} = 1 - \left(\frac{1}{2n}\sum d_i^2\right)\frac{1}{V(U)}$$

$$\frac{6\Sigma d_i^2}{\quad}$$

একে বলা হয় স্পীয়ারম্যানের (Spearman) মানক্রমিক সহগাঙ্ক। যদি দুই প্রস্থ অনুক্রম মানের মধ্যে পূর্ণ মিল থাকে, তবে প্রত্যেক $i=1, \ldots, n$-এর জন্যে $u_i=v_i$ ও ফলে $d_i=0$ অর্থাৎ $\Sigma d_i^2=0$ অর্থাৎ $r_{uv}=R_{AB}=1$ হবে। পক্ষান্তরে, যদি তাদের মধ্যে পূর্ণ অমিল থাকে, তবে $u_i=n-v_i+1$ হবে, অর্থাৎ $d_i=u_i-v_i=n-2v_i+1$ হবে এবং $\Sigma d_i^2=n(n+1)^2-2n(n+1)^2$ $+4\dfrac{n(n+1)(2n+1)}{6}=\dfrac{n(n^2-1)}{3}$. সুতরাং এক্ষেত্রে $R_{AB}=r_{uv}=-1$ হবে।

ওপরে মানক্রমিক সহগতির যে সংজ্ঞা ও সহগাঙ্কের সূত্র দেওয়া হ'ল তাতে ধরা হয়েছে যে উভয় চরিত্রানুযায়ীই প্রত্যেকটি ব্যষ্টির অনুক্রমমান পরস্পর পৃথক্। কিন্তু কখনও কখনও এমন হতে পারে যে, একাধিক ব্যষ্টি ঠিক সমপরিমাণে কোন বৈশিষ্ট্যের অধিকারী। সেক্ষেত্রে ঐ সকল ব্যষ্টির প্রত্যেককে একই অনুক্রমমান আরোপ করা উচিত। এরকম হলে বলা হয় যে, ঐ ব্যষ্টিগুলির মধ্যে ঐ চরিত্রবৈশিষ্ট্য অধিকারের ব্যাপারে সমতা বা সমানুক্রম (tie) সৃষ্টি হয়েছে। যদি দেখা যায় যে, $k_1$ সংখ্যক ব্যষ্টি সর্বাধিক মাত্রায় $A$ চরিত্র-বৈশিষ্ট্যের অধিকারী হয়, তবে তাদের প্রত্যেককে $\dfrac{1+2+\cdots+k_1}{k_1}=\dfrac{k_1+1}{2}$ অনুক্রমমান আরোপ করা দরকার (এখানে $k_1$ হচ্ছে $1, \ldots, n$ সংখ্যা-কটির যে কোন একটি)। তেমনি যদি ঠিক তৎপরবর্তী স্বল্পতর মাত্রার অধিকারী হয় $k_2$ সংখ্যক ব্যষ্টি, তবে ঐ $k_2$ সংখ্যক ব্যষ্টির প্রত্যেককে

$$\frac{(k_1+1)+(k_1+2)+\cdots+(k_1+k_2)}{k_2}=k_1+\frac{k_2+1}{2}$$

এই অনুক্রমান আরোপ করা দরকার। আবার ঠিক তৎপরবর্তী স্বল্পতর মাত্রার অধিকারী সংখ্যা যদি হয় $k_3$ তবে তাদের প্রত্যেককে

$$\frac{(k_1+k_2+1)+(k_1+k_2+2)+\cdots+(k_1+k_2+k_3)}{k_3}$$

$$=k_1+k_2+\frac{k_3+1}{2}$$

—এই অনুক্রমমান আরোপ করতে হবে ($k_3=1, 2, \ldots$). এখানে $k_1, k_2, k_3, \ldots.$ যদি 1-এর চেয়ে বড় হয়, তবে এই অনুক্রম মানগুলিকে বলা হয় সমানুক্রম মান (tied ranks) এবং $k_1, k_2, k_3, \cdots$-কে বলে সমানুক্রম দৈর্ঘ্য (tie-length). অন্যথায় এদেরকে অসমানুক্রম মান বা সংক্ষেপে শুধু অনুক্রম

মান বলে। এখন, মনে কর $A$-অনুযায়ী $n$ সংখ্যক ব্যষ্টিকে অনুক্রম মান আরোপ করলে $t$ সংখ্যক সমানুক্রম আছে ও $k_1, k_2, \ldots, k_i, \ldots, k_t$ হচ্ছে যথাক্রমে তাদের দৈর্ঘ্য এবং বাকী $n - \sum_{i=1}^{t} k_i$ সংখ্যক ব্যষ্টির প্রত্যেকের পৃথক্ পৃথক্ অনুক্রম মান রয়েছে। এক্ষেত্রে সবগুলি অনুক্রম মানের গড় ও ভেদমান কত হবে দেখা যাক। মনে কর $k_i$ সংখ্যক ব্যষ্টির প্রত্যেকের অনুক্রম মান হচ্ছে $\frac{(l+1)+\cdots+(l+k_i)}{k_i} = l + \frac{k_i+1}{2}$. তাহলে, এই কটি অনুক্রম মানের গড় হচ্ছে $\left(l + \frac{k_i+1}{2}\right)\frac{k_i}{k_i} = l + \frac{k_i+1}{2}$. আবার এদের যদি পৃথক্ পৃথক্ এবং পরপর মানক্রম (rank) হ'ত তাহলে সেই অনুক্রম মানগুলি হ'ত $l+1$, $l+2, \ldots, l+k_i$ এবং তাদের গড় হ'ত $\frac{(l+1)+\cdots+(l+k_i)}{k_i} = l + \frac{k_i+1}{2}$. কাজেই সমানুক্রম থাকার জন্যে অনুক্রম মানের গড়ে কোন পরিবর্তন হয় না অর্থাৎ সব কটি অনুক্রম মানের গড় এক্ষেত্রেও $\frac{n+1}{2}$-ই থাকবে। কিন্তু এই দুই জাতীয় অনুক্রম মানের ভেদমানের কথা বিবেচনা করতে গিয়ে দেখা যায় যে, সমানুক্রমের ক্ষেত্রে $\left(l + \frac{k_i+1}{2}\right)$ এই অনুক্রম মানগুলির বর্গসমষ্টি হচ্ছে $R_i = k_i\left[l + \frac{k_i+1}{2}\right]^2 = l^2k_i + lk_i(k_i+1) + \frac{1}{4}k_i(k_i+1)^2$. কিন্তু, যদি অনুক্রম মানগুলি পৃথক্ পৃথক্ অর্থাৎ $(l+1), (l+2), \ldots, (l+k_i)$ হয়, তবে তাদের বর্গসমষ্টি হবে

$$S_i = (l+1)^2 + \cdots + (l+k_i)^2 = l^2k_i + lk_i(k_i+1) + \tfrac{1}{6}k_i(k_i+1)(2k_i+1).$$

কাজেই তাদের পার্থক্য হচ্ছে $D_i = R_i - S_i = \frac{k_i(k_i^2-1)}{12}$. সুতরাং যেহেতু দুই প্রস্থ অনুক্রম মানের গড় অপরিবর্তিত, তাদের ভেদমানের পার্থক্য হবে $\frac{k_i(k_i^2-1)}{12n}$-এর সমান। সুতরাং মোট $n$-সংখ্যক ব্যষ্টির ভেদমান হবে সমানুক্রম না থাকলে যত ভেদমান হ'ত তার থেকে $\frac{1}{12n}\sum_{i=1}^{t} k_i(k_i^2-1)$

পরিমাণ কম অর্থাৎ ভেদমান $V(U)=\frac{n^2-1}{12}-\frac{1}{12n}\sum_{i=1}^{t}k_i(k_i^2-1)$. তেমনি যদি $B$-চরিত্রবৈশিষ্ট্য অনুযায়ী $m$ সংখ্যক সমানুক্রম মান থাকে ও তাদের দৈর্ঘ্য হয় যথাক্রমে $k'_1, k'_2, \ldots, k'_i, \ldots, k'_m$ এবং অন্যান্য $\left(n-\sum_{i=1}^{m}k'_i\right)$ সংখ্যক ব্যষ্টির অনুক্রম মান পৃথক্ পৃথক্ হয়, তবে তাদের গড় ও ভেদমান হবে যথাক্রমে $\frac{n+1}{2}$ ও $\frac{n^2-1}{12}-\frac{1}{12n}\sum_{i=1}^{m}k'_i(k'^2_i-1)$. অবশ্য, $d_i=u_i-v_i$-এর মান সমানুক্রম মান থাকার জন্যে পরিবর্তিত হবে না। তাই cov $(U, V)=\frac{V(U)}{2}$

$$+\frac{V(V)}{2}-\frac{1}{2n}\sum d_i^2=\frac{n^2-1}{12}-\frac{1}{12n}\cdot\frac{1}{2}\sum_{i=1}^{t}k_i(k_i^2\times 1)$$

$$-\frac{1}{12n}\cdot\frac{1}{2}\cdot\sum_{i=1}^{m}k'(k'^2_i-1)-\frac{1}{2n}\sum d_i^2.$$

$$\text{সুতরাং } R_{AB}=\frac{\frac{n^2-1}{12}-\frac{T_u+T_v}{2}-\frac{1}{2n}\sum d_i^2}{\sqrt{\left(\frac{n^2-1}{12}-T_u\right)\left(\frac{n^2-1}{12}-T_v\right)}};$$

$$\text{এখানে } T_u=\frac{1}{12n}\sum k_i(k_i^2-1) \text{ ও } T_v=\frac{1}{12n}\sum_{i=1}^{m}k'_i\,(k'^2_i-1)$$

লেখা হয়েছে। এক্ষেত্রেও অবশ্য যদি $A$ ও $B$ অনুযায়ী দুই প্রস্থ অনুক্রম মানের মধ্যে সম্পূর্ণ মিল থাকে, তাহলে $u_i=v_i$ হবে এবং তার ফলে $t=m$, $k_i=k'_i\ (i=1,\ldots t)$,

$$\sum d_i^2=0 \text{ এবং } T_u=\frac{1}{12n}\sum k_i\,(k_i^2-1)=T_v \text{ হবে}$$

$$\text{সুতরাং এক্ষেত্রে } R_{AB}=r_{uv}=\frac{\frac{n^2-1}{12}-T_u}{\frac{n^2-1}{12}-T_u}=1 \text{ হবে।}$$

মানক্রমিক সহগতি নির্ণয়ের জন্যে আরও একটি সহগাঙ্ক অনেক সময় ব্যবহার করা হয়। তাকে বলে কেণ্ডালের মানক্রমিক সহগাঙ্ক (Kendall's rank correlation coefficient). এখানে আগের মতোই $A$ ও $B$ চরিত্রবৈশিষ্ট্যানুযায়ী $n$-সংখ্যক ব্যষ্টিকে অনুক্রম মান $U$ ও $V$ আরোপ করা হয়। তারপর $i$-তম ও $j$-তম ব্যষ্টিদ্বয়ের জন্যে যদি দেখা যায় যে, $u_i > u_j$ হলে $v_i > v_j$ হয়, তাহলে $(i, j)$-তম ব্যষ্টিযুগ্মকে $+1$ ও পক্ষান্তরে $u_i > u_j$ হলে যদি $v_i < v_j$ হয়, তাহলে তাকে $-1$ এই সংখ্যাটি আরোপ করা হয়। ঠিক এই ব্যাপারটি $\binom{n}{2}$ সংখ্যক ব্যষ্টিযুগ্মের প্রত্যেকের জন্যেই করা হয় এবং এইভাবে প্রাপ্ত সংখ্যাগুলির যোগফল নেওয়া হয়। এই যোগফলকে বলা হয় মোটসংখ্যা বা পূর্ণমান। স্পষ্টতঃই এই মোট সংখ্যার সর্বোচ্চ মান হতে পারে $\binom{n}{2}$. এই সর্বোচ্চ মান পাওয়া যাবে যদি প্রত্যেকটি ব্যষ্টিযুগ্ম $(i, j)$-এর জন্যেই প্রাপ্ত সংখ্যা হয় $+1$ অর্থাৎ যদি $A$ অনুযায়ী মানক্রমগুলি বাড়ার ( কমার ) সঙ্গে সঙ্গে $B$ অনুযায়ী নির্ণীত প্রত্যেকটি মানক্রমই বাড়তে ( কমতে ) থাকে। এখন,

$$\tau = \frac{n\text{-সংখ্যক ব্যষ্টির জন্যে প্রাপ্ত মোট নম্বর}}{n\text{-সংখ্যক ব্যষ্টির জন্যে সর্বোচ্চ মোট নম্বর} = \binom{n}{2}},—$$

এই অনুপাতটিকে $A$ ও $B$ এই দুটি চরিত্রবৈশিষ্ট্যের সহগতির একটি মাপক হিসেবে নেওয়া হয়। এই সূত্রটি প্রথম নির্দেশ করেন মরিস কেণ্ডাল (M. G. Kendall). এই জন্যে একে কেণ্ডালের মানক্রমিক সহগাঙ্ক বলে। একে অনেক সময় সংক্ষেপে কেণ্ডালের $\tau$ বলা হয়।

কেণ্ডালের $\tau$-এর মান নির্ণয়ের একটি সহজতর পন্থা আছে। মনে কর $A$ অনুযায়ী অনুক্রম মান $u_i$ গুলিকে সব মানের স্বাভাবিক ঊর্ধ্ব ক্রমানুসারে (natural order) অর্থাৎ $(1, 2, \ldots, n)$ পর্যায়ক্রমে লেখা হ'ল। এখন মনে কর $\theta_i$ হচ্ছে $B$ অনুযায়ী সেই ব্যষ্টির মানক্রম $A$-চরিত্র অনুযায়ী যার মানক্রম হচ্ছে $i$ $(i=1, \ldots, n)$ অর্থাৎ $(\theta_1, \ldots \theta_i, \ldots, \theta_n)$ হচ্ছে $(v_1, \ldots, v_i, \ldots, v_n)$-এর একটি বিন্যাস। এখন, ধর $f_i$ হচ্ছে $(\theta_{i+1}, \theta_{i+2}, \ldots, \theta_n)$-এর মধ্যে মোট যতগুলি মানক্রম $\theta_i$-এর চেয়ে বড় ততসংখ্যা এবং $P = \sum_{i=1}^{n} f_i$. তাহলে, স্পষ্টতঃই,

$P$ হচ্ছে মোট যতগুলি ব্যষ্টিযুগ্মের জন্যে $u_i$ ও $v_i$ একই সঙ্গে বাড়ছে বা কমছে অর্থাৎ মোট যতগুলি ব্যষ্টিযুগ্মের জন্যে নম্বর হবে $+1$. মনে কর $Q=\binom{n}{2}-P$. তাহলে, মোট $Q$ সংখ্যক ব্যষ্টিযুগ্মের জন্যে প্রাপ্তসংখ্যা হবে $-1$ অর্থাৎ তাদের জন্যে $u_i$ বাড়লে $v_i$ কমবে। তাহলে, মোট সংখ্যা বা পূর্ণমান হবে $P-Q$. সুতরাং, সংজ্ঞানুযায়ী

$$\tau=\frac{P-Q}{\binom{n}{2}}=\frac{P-Q}{P+Q}=\frac{2P}{\binom{n}{2}}-1=1-\frac{2Q}{\binom{n}{2}}.$$

এছাড়া, $\tau$-এর মান নির্ণয়ের আরও একটি উপায় আছে। মনে কর,

$$a_{ij}=\begin{cases}+1 & \text{যদি } i<j \text{ ও } u_i<u_j \text{ হয়}\\ -1 & \text{যদি } i<j \text{ ও } u_i>u_j \text{ হয়}\end{cases}$$

$$\text{এবং}\quad b_{ij}=\begin{cases}+1 & \text{যদি } i<j \text{ ও } u_i>u_j \text{ হয়}\\ -1 & \text{যদি } i<j \text{ ও } v_i>v_j \text{ হয়।}\end{cases}$$

$$\text{তাহলে,}\quad \tau=\frac{\sum\sum_{i<j} a_{ij}b_{ij}}{\sqrt{\sum\sum_{i<j} a_{ij}^2}\cdot\sqrt{\sum\sum_{i<j} b_{ij}^2}}$$

$$\text{লক্ষণীয় যে,}\quad \sum\sum_{i<j} a_{ij}^2=\binom{n}{2}=\frac{n(n-1)}{2}=\sum\sum_{i<j} b_{ij}^2.$$

এতক্ষণ যা বলা হ'ল তা খাটবে যদি কোন সমানুক্রম মান না থাকে। যদি সমানুক্রম মান থাকে, তাহলে $\tau$-এর সংজ্ঞায় কিছু পরিবর্তন হবে। আমরা $i<j$ নিয়ে লিখব

$$a_{ij}=\begin{cases}+1 & \text{যদি } u_i<u_j \text{ হয়}\\ 0 & \text{,, } u_i=u_j \text{ ,,}\\ -1 & \text{,, } u_i>u_j \text{ ,,}\end{cases}$$

$$\text{এবং}\quad b_{ij}=\begin{cases}+1 & \text{,, } v_i<v_j \text{ ,,}\\ 0 & \text{,, } v_i=v_j \text{ ,,}\\ -1 & \text{,, } v_i>v_j \text{ ,,.}\end{cases}$$

তাহলে, $A$-এর জন্যে যদি একটি $k_1$ দৈর্ঘ্যবিশিষ্ট সমানুক্রম মান থাকে, তবে $\frac{k_1(k_1-1)}{2}$ সংখ্যক ব্যষ্টিযুগ্মের জন্যে $a_{ij}=0$ হবে এবং যদি $A$-এর জন্যে $t$-সংখ্যক সমানুক্রম মান থাকে এবং তাদের দৈর্ঘ্য যথাক্রমে $k_1, \cdots, k_i, \ldots k_t$ হয়, তবে মোট $\sum_{i=1}^{t} \frac{k_i(k_i-1)}{2}$ সংখ্যক ব্যষ্টিযুগ্মের জন্যে $a_{ij}=0$ হবে। কাজেই এক্ষেত্রে

$$\sum_{i<j}\sum a_{ij}^2 = \frac{n(n-1)}{2} - \frac{1}{2}\sum_{i=1}^{t} k_i(k_i-1).$$ তেমনি যদি $B$ অনুযায়ী মানক্রমে $m$ সংখ্যক সমানুক্রম মান থাকে এবং তাদের দৈর্ঘ্য হয় $k'_1, \ldots, k'_i, \ldots k'_m$, তাহলে মোট $\frac{1}{2}\sum_{i=1}^{m} k'_i(k'_i-1)$ সংখ্যক ব্যষ্টিযুগ্মের জন্যে $b_{ij}=0$ হবে। ফলে,

$$\sum_{i<j}\sum b_{ij}^2 = \frac{n(n-1)}{2} - \frac{1}{2}\sum_{i=1}^{m} k_i'(k'_i-1)$$ হবে। এদিকে মোট প্রাপ্ত সংখ্যা $\sum_{i<j}\sum a_{ij}\,b_{ij}$-এর মান অবশ্য ওপরের সংজ্ঞানুযায়ী সহজেই নির্ণেয়। কাজেই সমানুক্রম মানের অস্তিত্ব থাকলে কেণ্ডালের $\tau$ দাঁড়াবে নিম্নরূপ :

$$\tau = \frac{\sum_{i<j}\sum a_{ij}\,b_{jj}}{\sqrt{\left\{\frac{n(n-1)}{2} - \frac{1}{2}\sum k_i(k_i-1)\right\}}\sqrt{\left\{\frac{n(n-1)}{2} - \frac{1}{2}\sum k'_i(k'_i-1)\right\}}}$$

## 10.14 অন্তঃশ্রেণীক সহগতি (Intra class correlation) :

এতক্ষণ পর্যন্ত আমরা যে সমস্ত সহগতির আলোচনা করেছি তাতে সর্বদাই দুটি ক'রে পৃথক্ চল বিবেচিত হয়েছে ; যেমন (1) দৈহিক উচ্চতা ও ওজনের সহগতি, (2) ইংরেজী ও অঙ্কের নম্বরের সহগতি ইত্যাদি। এগুলিকে আন্তঃ-শ্রেণীক (inter class) সহগতি বলা যায়, কারণ যে দুটি চলের মানগুলি নিয়ে

সহগতি নির্ধারণ করা হয় সেগুলিকে দুটি পৃথক্ শ্রেণীভুক্ত রাশি ব'লে মনে করা যায়। কিন্তু অনেক সময় এমন প্রয়োজনের উদ্ভব হয় যেখানে একটিমাত্র চল কোন শ্রেণীভুক্ত বিভিন্ন সদস্য ব্যষ্টিগুলির মধ্যে কিভাবে সহগতিযুক্ত তা নির্ণয় করতে হয়। যেমন, মনে কর আমরা জানতে ইচ্ছুক হতে পারি বিভিন্ন পরিবারে দৈহিক উচ্চতা সম্পর্কে বিভিন্ন ভাইবোনের মধ্যে কী ধরনের এবং কতখানি সংস্রব রয়েছে। এখানে রাশিগুলিকে একই শ্রেণীভুক্ত মান ব'লে ধরতে পারি, কারণ তারা সব উচ্চতা-চলটিরই মান। কিন্তু তা সত্ত্বেও একটি কৌশলের সাহায্যে ঐ একশ্রেণীভুক্ত মানগুলি থেকে দুই শ্রেণীর মান নির্দিষ্ট ক'রে তাদেরকে দুটি ভিন্ন চলের মান ব'লে ধ'রে ঐ চল-দুটির মধ্যে সহগতি নির্ধারণ ক'রে মূল একটিমাত্র চলের পূর্বালোচিত সংস্রব নির্ণয় করা হয়। এজন্যে একে অন্তঃশ্রেণীক (Intra class) সহগতি বলে। এর সঙ্গে পার্থক্য দর্শাবার জন্যে পূর্বালোচিত সহগতিকে আন্তঃশ্রেণীক সহগতি বলা হয়, কারণ সেই সহগতি দুই শ্রেণীর মানের ( অর্থাৎ দুটি ভিন্ন চলের ) ভিত্তিতে নির্ণীত সহগতি।

মনে কর $p$ সংখ্যক বিভিন্ন পরিবার আছে এবং $i$-তম পরিবারে মোট $k_i$ সংখ্যক সদস্যভ্রাতা আছে $(i=1,\ldots,p)$ এবং $x_{ij}$ হচ্ছে $i$-তম পরিবারের $j$-তম ভ্রাতার সম্পর্কে নির্ণীত কোন চল X-এর মান ( দৃষ্টান্তস্বরূপ, দৈহিক উচ্চতার মান )। এখন, একটি সারণী এমনভাবে গঠন করা যাক যাতে দুটি স্তম্ভ (column) আছে, যার প্রথমটিতে গোড়ায় একাদিক্রমে প্রথম পরিবারের প্রথম ভ্রাতার $X$ মানগুলি ও তাদের পাশে পাশে দ্বিতীয় স্তম্ভে প্রথম পরিবারের বাকী $(k_1-1)$ সংখ্যক ভ্রাতার $X$ মানগুলি লেখা হ'ল এবং একইভাবে প্রথম পরিবারের বাকী $(k_1-1)$ ভ্রাতার $X$ মান প্রথম স্তম্ভে ও তাদের প্রত্যেকের পাশে পাশে দ্বিতীয় স্তম্ভে অন্য $(k_1-1)$ সংখ্যক ভ্রাতার $X$ মান লেখা হ'ল এবং এই ভাবে $p$ সংখ্যক পরিবারের প্রত্যেকটি ভ্রাতার জন্যে একইভাবে মানগুলি দুটি স্তম্ভে লিখে একটি সুষম (symmetrical) সারণী গঠন করা হ'ল যার উভয় স্তম্ভে একই রাশিগুচ্ছ লেখা হ'ল, অবশ্য ভিন্নতর বিন্যাসে। এখন এই সারণীর প্রথম স্তম্ভের মানগুলিকে একটি চল $U$-এর মান ও দ্বিতীয় স্তম্ভের মানগুলিকে অন্য একটি চল $V$-এর মান হিসেবে ধরে $U$ ও $V$-এর সহগতি নির্ণয় করা যায়। এই সহগতিকে $X$ চলের অন্তঃশ্রেণীক সহগতি বলে। এই সহগতি নির্ণয় করা হয় $U$ ও $V$-এর মধ্যে সহগাঙ্ক $r_{uv}$ নির্ণয় ক'রে। একে বলে $X$-এর অন্তঃশ্রেণীক সহগাঙ্ক (Intra class correlation coefficient).

পূর্বোক্ত সারণীটির চেহারা তাহলে দাঁড়াবে নিম্নরূপ :—

### সারণী 10.5

### অন্তঃশ্রেণীক সহগাঙ্ক নির্ণয়ার্থে ব্যবহার্য সুষম সারণী

| পরিবারের ক্রমিক → সংখ্যা | 1 | | 2 | | ………… | p | |
|---|---|---|---|---|---|---|---|
| | $U$ | $V$ | $U$ | $V$ | | $U$ | $V$ |
| | $x_{11}$ | $x_{12}$ | $x_{21}$ | $x_{22}$ | | $x_{p1}$ | $x_{p2}$ |
| | $x_{11}$ | $x_{13}$ | $x_{21}$ | $x_{23}$ | | $x_{p1}$ | $x_{p3}$ |
| | . | . | . | . | | . | |
| | . | . | . | . | | . | |
| | $x_{11}$ | $x_{1k_1}$ | $x_{21}$ | $x_{2k_2}$ | | $x_{p1}$ | $x_{pk_p}$ |
| | …… | | …… | | | …… | |
| | $x_{12}$ | $x_{11}$ | $x_{22}$ | $x_{21}$ | | $x_{p2}$ | $x_{p1}$ |
| | $x_{12}$ | $x_{13}$ | $x_{22}$ | $x_{23}$ | | $x_{p2}$ | $x_{p3}$ |
| | . | . | . | | | . | |
| | . | . | . | | | . | |
| | $x_{12}$ | $x_{1k_1}$ | $x_{22}$ | $x_{2k_2}$ | | $x_{p2}$ | $x_{pk_p}$ |
| | …… | | …… | | | …… | |
| | . | | . | | | . | |
| | . | | . | | | . | |
| | $x_{1k_1}$ | $x_{11}$ | $x_{2k_2}$ | $x_{21}$ | | $x_{pk_p}$ | $x_{p1}$ |
| | $x_{1k_1}$ | $x_{12}$ | $x_{2k_2}$ | $x_{22}$ | | $x_{pk_p}$ | $x_{p2}$ |
| | . | | . | | | . | |
| | . | | $x_{2k_2}$ | $x_{2\overline{k_2-2}}$ | | . | |
| | $x_{1k_1}$ | $x_{1\overline{k_1-1}}$ | $x_{2k_2}$ | $x_{2\overline{k_2-1}}$ | | $x_{pk_p}$ | $x_{p\overline{k_p-1}}$ |

তাহলে, $U$ ও $V$ প্রত্যেকেরই মোট $N=\sum_{j=1}^{p} k_i(k_i-1)$ সংখ্যক মান রয়েছে এবং স্পষ্টতঃই তাদের উভয়েরই সামগ্রিক গড় ও ভেদমান সমমান-বিশিষ্ট এবং তারা হ'ল যথাক্রমে

$$\bar{U}=\bar{V}=\frac{1}{N}\sum_{i=1}^{p}(k_i-1)\sum_{j=1}^{k_i}x_{ij}=\bar{x}_0 \text{ ( এটি সাধারণ গড় নয় ),}$$

$$V(U)=V(V)=S_U{}^2=S_V{}^2=\frac{1}{N}\sum_{i=1}^{p}(k_i-1)\sum_{j=1}^{k_i}(x_{ij}-\bar{x}_0)^2.$$

[ এটিও সাধারণ ভেদমান নয় ].

কিন্তু $$\operatorname{cov}(U,V)=\frac{1}{N}\sum_{i=1}^{p}\sum_{\substack{j=1\\ j\neq j'}}^{k_i}\sum_{j'=1}^{k_i}(x_{ij}-\bar{x}_0)(x_{ij'}-\bar{x}_0).$$

এখন, $$\sum_{\substack{j=1\\ j\neq j'}}^{k_i}\sum_{j'=1}^{k_i}(x_{ij}-\bar{x}_0)(x_{ij'}-\bar{x}_0)$$

$$=\sum_{j=1}^{k_i}\sum_{j'=1}^{k_i}(x_{ij}-\bar{x}_0)(x_{ij'}-\bar{x}_0)-\sum_{j=1}^{k_i}(x_{ij}-\bar{x}_0)^2$$

$$=\sum_{j=1}^{k_i}(x_{ij}-\bar{x}_0)\sum_{j'=1}^{k_i}(x_{ij'}-\bar{x}_0)-\sum_{j=1}^{k_i}(x_{ij}-\bar{x}_0)^2$$

$$=\left\{\sum_{j=1}^{k_i}(x_{ij}-\bar{x}_0)\right\}^2-\sum_{j=1}^{k_i}(x_{ij}-\bar{x}_0)^2$$

$$=k_i{}^2(\bar{x}_i-\bar{x}_0)^2-\sum_{j=1}^{k_i}(x_{ij}-\bar{x}_0)^2.$$

এখানে, $\bar{x}_i = \frac{1}{k_i}\sum_{j=1}^{k_i} x_{ij}$ হচ্ছে $i$-তম পরিবারভুক্ত ভ্রাতাদের জন্যে $X$-এর অর্থাৎ $U$ ও $V$-এর গড়। তাহলে,

$$\text{cov}\,(U, V) = \frac{1}{N}\Big[\sum_{i=1}^{k_i} k_i^2(\bar{x}_i - \bar{x}_0)^2 - \sum_{i=1}^{p}\sum_{j=1}^{k_i}(x_{ij} - \bar{x}_0)^2\Big]$$

এবং $r(U, V) \quad \dfrac{\text{cov}\,(U, V)}{\sqrt{V(U)}\cdot\sqrt{V(V)}}$,

অর্থাৎ $X$-এর অন্তঃশ্রেণীক সহগাঙ্ক $r_I$ হচ্ছে

$$r_I = r(U, V) = \frac{\sum_{i=1}^{p} k_i^2(\bar{x}_i - \bar{x}_0)^2 - \sum_{i=1}^{p}\sum_{j=1}^{k_i}(x_{ij} - \bar{x}_0)}{\sum_{i=1}^{p}(k_i - 1)\sum_{j=1}^{k_i}(x_{ij} - \bar{x}_0)^2}$$

যদি প্রত্যেক পরিবারে ভ্রাতৃসংখ্যা সমান হয় অর্থাৎ যদি প্রত্যেক $i = 1, \ldots, p$-এর জন্যে $k_i = k$ হয়, তাহলে পাব

$$N = pk\,(k-1),\ \bar{x}_i = \frac{1}{k}\sum_{j=1}^{k} x_{ij},\ \bar{x}_0 = \frac{k-1}{pk(k-1)}\sum_{i=1}^{p}\sum_{j=1}^{k} x_{ij}$$

$$= \frac{1}{pk}\sum_{i=1}^{p}\sum_{j=1}^{k} x_{ij}$$

$$\text{Cov}\ \ (U, V) = \frac{1}{pk(k-1)}\Big[k^2\sum_{i=1}^{p}(\bar{x}_i - \bar{x}_0)^2 - \sum_{i=1}^{p}\sum_{j=1}^{k}(x_{ij} - \bar{x}_0)^2\Big]$$

$$V(U) = V(V) = S_u^2 = S_v^2 = \frac{(k-1)}{\{pk(k-1)\}}\sum_{i=1}^{p}\sum_{j=1}^{k}(x_{ij} - \bar{x}_0)^2$$

$$= \frac{1}{pk}\sum_{i=1}^{p}\sum_{j=1}^{k}(x_{ij} - \bar{x}_0)^2 = S^2$$

এবং $$r_I = \frac{\text{Cov}(U, V)}{\sqrt{V(U)\,V(V)}}$$

$$= \frac{\frac{k}{p(k-1)}\sum_{i=1}^{p}(\bar{x}_i - \bar{x}_o)^2 - \frac{1}{pk(k-1)}\sum_{i=1}^{p}\sum_{j=1}^{k}(x_{ij} - \bar{x}_o)^2}{\frac{1}{pk}\sum_{i=1}^{p}\sum_{j=1}^{k}(x_{ij} - \bar{x}_o)^2}$$

$$= \frac{k}{k-1}\left[\frac{1}{p}\sum_{i=1}^{p}(\bar{x}_i - \bar{x}_o)^2\right]\frac{1}{S^2} - \frac{1}{(k-1)}$$

$$= \frac{k}{(k-1)}\cdot\frac{S_m^2}{S^2} - \frac{1}{(k-1)} = \frac{1}{(k-1)}\left[k\frac{S_m^2}{S^2} - 1\right];$$

এখানে আমরা লিখেছি $S_m^2 = \frac{1}{p}\sum_{i=1}^{p}(\bar{x}_i - \bar{x}_o)^2$

$= p$ সংখ্যক পরিবারভুক্ত গড় মানগুলির ভেদমান।

এখন, আমরা দেখতে পারি যে,

$$\sum_{i=1}^{p}\sum_{j=1}^{k}(x_{ij} - \bar{x}_o)^2 = \sum_{i=1}^{p}\sum_{j=1}^{k}\{(x_{ij} - \bar{x}_i) + (\bar{x}_i - \bar{x}_o)\}^2$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{k}(x_{ij} - \bar{x}_i)^2 + k\sum_{i=1}^{p}(\bar{x}_i - \bar{x}_o)^2$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{k}(x_{ij} - \bar{x}_i)^2 + pkS_m^2.$$

সুতরাং $$S^2 = \frac{1}{pk}\sum_{i=1}^{p}\sum_{j=1}^{k}(x_{ij} - \bar{x}_o)^2$$

$$= S_m^2 + \frac{1}{pk}\sum_{i=1}^{p}\sum_{j=1}^{k}(x_{ij} - \bar{x}_i)^2.$$

এখন, $\frac{1}{pk}\sum_{i=1}^{p}\sum_{j=1}^{k}(x_{ij} - \bar{x}_i)^2 = S_w^2$-কে বলা যেতে পারে মানগুলির পরিবারমধ্যস্থ ভেদমান। কাজেই $S_w^2 > 0$ এবং $S^2 > S_m^2 > 0$.

ফলে, $r_I = \frac{1}{(k-1)}\left[k\frac{S_m^2}{S^2} - 1\right] < \frac{1}{(k-1)}(k-1) = 1$

এবং $r_I > \frac{-1}{k-1}$.

এখন, $\sum_i \sum_j (x_{ij} - \bar{x}_i)^2 = 0$ অর্থাৎ প্রত্যেক $j = 1, \cdots, n$ এর জন্যে $x_{ij} = \bar{x}_i (i = 1, \cdots, k)$ হলে অর্থাৎ প্রত্যেক পরিবারের জন্যেই পৃথক্ পৃথক্ ভাবে মানগুলি পরিবার-গড়মানের সমান হলে $S_m^2$ তার সর্বোচ্চ মান $S^2$-এর সমান হয় এবং সেক্ষেত্রে $r_I$-এর মানও সর্বোচ্চ $(=1)$ হয়। পক্ষান্তরে, $\Sigma(\bar{x}_i - \bar{x})^2 = 0$ অর্থাৎ পরিবার-গড়মানগুলি পরস্পর সমান হলে $S_m^2$-এর মান সর্বনিম্ন $(=0)$ হয় ও তখন $r_I$-এর মানও সর্বনিম্ন $\left(= \frac{-1}{k-1}\right)$ হয়। ফলে, $S_m^2$ তার $[0, S^2]$-এর অন্তর্বর্তী মানগুলি গ্রহণ করার সঙ্গে সঙ্গে $r_I$ ও তার $\left[-\frac{1}{k-1}, 1\right]$-এর অন্তর্বর্তী বিভিন্ন মান ধারণ করে। কাজেই সংক্ষেপে বলা যায় যে $r_I$-কে অবেক্ষিত মানগুলির সামগ্রিক প্রভেদের যতটুকু পারিবারিক গড়মানগুলির প্রভেদের সাহায্যে ব্যাখ্যাত হয় তার একটি মাপক হিসেবে গণ্য করা যায়।

## অনুশীলনী

10.1 মনে কর, দুটি সম্ভাবনা চল $X$ ও $Y$ উভয়েই কেবলমাত্র দুটি ক'রে পৃথক্ মান গ্রহণ করে। দেখাও যে এদের সহগাঙ্ক যদি শূন্য হয়, তবে তারা পরস্পর নির্ভরতাশূন্য হবে।

10.2 $y = -1{\cdot}32x + {\cdot}26$

এবং $x = {\cdot}69y - 1{\cdot}34$

—এই সমীকরণ দুটি কি $x$-এর ওপর $y$-এর এবং $y$-এর ওপর $x$-এর নির্ভরণ রেখা নির্দেশ করতে পারে ?

[ উত্তর : না।

আভাস : $b_{yx} = r\frac{s_y}{s_x}$ এবং $b_{xy} = r\frac{s_x}{s_y}$ ; অতএব $b_{yx}$ ও $b_{xy}$ সমচিহ্নযুক্ত হবে ]

10.3 $y = 3{\cdot}02x + {\cdot}61$

ও $x = {\cdot}13y - {\cdot}24$

যদি $x(y)$-এর ওপর $y(x)$-এর নির্ভরণ রেখা নির্দেশ করে, তাহলে $\bar{x}, \bar{y}$ এবং $r$-এর মান নির্ণয় কর। $s_x$ ও $s_y$-এর মানও কি প্রদত্ত তথ্য থেকে নির্ণয় করা যাবে ?

[ উত্তর : $\bar{x} = {\cdot}53, \bar{y} = 2{\cdot}21, r = {\cdot}627.$ না ]

10.4 দেখাও যে, $f(x, y) = 3x^2 - 8xy + 6y^2, 0 < x < 1, 0 < y < 1,$ অপেক্ষককে একটি দ্বিচল সম্ভাবনা বিভাজনের সম্ভাবনা-ঘনত্ব-অপেক্ষক হিসেবে ধরা যায়। এক্ষেত্রে আনুসঙ্গিক প্রান্তীয় এবং সর্তাধীন সম্ভাবনা বিভাজনের ঘনত্ব-অপেক্ষকগুলি নির্ণয় কর।

[ উত্তর : $g(x) = 3x^2 - 4x + 2$

$h(y) = 1 - 4y + 6y^2$ ]

10.5 যদি $V(X) = V(Y)$ হয়, তাহলে দেখাও যে,

$$\rho(X + Y, X - Y) = 0.$$

10.6 $X$ এবং $S - X$ চল-দুটির সহগাঙ্ক কত ?

[ উত্তর : $-1$ ]

10.7 $X$ ও $Y$ হচ্ছে দুটি চল। তাদের পরিঘাত-গুণনজাত (product-moment) সহগাঙ্ক কত ? এই সহগাঙ্কের কয়েকটি গুণধর্ম বর্ণনা কর।

[ আভাস : $$r = \frac{\frac{1}{n}\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\frac{1}{n}\sum(x - \bar{x})^2}\sqrt{\frac{1}{n}\sum(y - \bar{y})^2}}$$-কে

পরিঘাত-গুণনজাত সহগাঙ্ক বলা হয় কারণ এর লবে ব্যবহৃত

$$\frac{1}{n}\sum(x - \bar{x})(y - \bar{y})$$-কে

$X$ ও $Y$-এর গুণনজাত প্রথম পরিঘাত বলা যায় কারণ এটি হচ্ছে $X$ ও $Y$-এর যৌথ বিভাজনের প্রথম গড়কেন্দ্রিক পরিঘাত $\mu_{11}$ ; কারণ এতে $(x - \bar{x})$ এবং $(y - \bar{y})$ উভয়েরই সূচক নেওয়া হয়েছে 1 এবং এদেরকে গুণ ক'রে তার গড় নেওয়া হয়েছে ]।

10.8 $X$-এর ওপর $Y$-এর নির্ভরণ বলতে কী বোঝায় ? লঘিষ্ঠ বর্গনীতি কী ? নর্ম্যাল সমীকরণ কাকে বলে ?

10.9 দুটি চল $X$ ও $Y$-এর সহগতি নির্ণয়ে সহগাঙ্ক $r$-এর প্রয়োগের সার্থকতা ও ব্যর্থতা আলোচনা কর।

10.10 নীচের সারণীতে কয়েকটি পরিবারে পিতা ও তাঁদের জ্যেষ্ঠপুত্রের দৈহিক ওজনের হিসাব দেওয়া আছে। পিতা ও পুত্রের ওজনের মধ্যে কোন সহগতি আছে কিনা মানক্রমিক সহগাঙ্ক নির্ণয় ক'রে সে সম্পর্কে আলোচনা কর।

**সারণী 10.9**

| পরিবার ক্রমিক সংখ্যা | পিতার ওজন ( গ্রাম ) | জ্যেষ্ঠপুত্রের ওজন ( গ্রাম ) |
|---|---|---|
| 1 | 53816 | 47943 |
| 2 | 76019 | 59112 |
| 3 | 59013 | 64032 |
| 4 | 57335 | 51765 |
| 5 | 66018 | 71239 |
| 6 | 55191 | 57034 |
| 7 | 50589 | 51314 |
| 8 | 57351 | 53469 |

[ উত্তর :

10.11 নীচের সারণীতে দুটি পরস্পর সম্পর্কযুক্ত বিষয় $A$ এবং $B$-তে 1000 জন ছাত্রছাত্রী কোন পরীক্ষায় যত নম্বর পেয়েছে তার বিস্তারিত বিভাজন দেওয়া আছে।

(*a*) $A$ এবং $B$-তে প্রাপ্ত নম্বরের সহগাঙ্ক নির্ণয় কর।

(*b*) $x$-এর ওপর $y$-এর ঋজুরৈখিক নির্ভরণ সমীকরণ নিরূপণ কর।

(*c*) উপযুক্ত নির্ভরণ সমীকরণ ব্যবহার ক'রে হিসেব কর কেউ যদি $A$-তে 52 নম্বর পায় তবে $B$-তে তার কত নম্বর পাওয়ার প্রত্যাশা হতে পারে?

(*d*) $x$-এর ওপর $y$-এর সহগতি অনুপাত নির্ণয় কর।

10.12 দেখাও যে, $\text{cov}(X+Y, X) = V(X) + \text{cov}(X, Y)$.

সারণী 10.10

| $y$ ( $B$-তে প্রাপ্ত নম্বর ) | $x$ ( $A$-তে প্রাপ্ত নম্বর ) | | | | | | |
|---|---|---|---|---|---|---|---|
| | 10-19 | 20-29 | 30-39 | 40-49 | 50-59 | 60-69 | 70-79 |
| 10-19 | 2 | 3 | 3 | 1 | — | — | — |
| 20-29 | 1 | 7 | 28 | 16 | 2 | 1 | — |
| 30-39 | 1 | 13 | 89 | 127 | 43 | 2 | — |
| 40-49 | — | 4 | 84 | 205 | 125 | 19 | 1 |
| 50-59 | — | — | 14 | 73 | 78 | 26 | 2 |
| 60-69 | — | — | — | 4 | 11 | 11 | 2 |
| 70-79 | — | — | — | — | 1 | 1 | — |

## নির্দেশিকা


1. Ezekiel, M and Fou, K. A. *Methods of Correlation and Regression Analysis.* John Wiley, 1959.

2. Goon, A M. ; Gupta, M.K. and Dasgupta, B. *Fundamentals of Statistics* ; Vol. I. The World Press Pvt. Ltd., 1970.

3. Goulden, C.H. *Methods of Statistical Analysis.* Asia Publishing House, 1959.

4. Keeping, E.S. and Kenney, J.F. *Mathematics of Statistics*, Part I, Van Nostrand, 1954.

5. Yule, G.U. and Kendall, M.C. *An Introduction to the Theory of Statistics.* Charles Griffin, 1950.

# 11 সহগতি ও নির্ভরণ : 2
## ( Correlation and Regression : 2 )

**11.1 বহুচল বিভাজন :** আগে যেমন দ্বিচল বিভাজন নিয়ে আলোচনা করেছি, তেমনি অনেক সময় একই সঙ্গে অনেকগুলি চলের যৌথ বিভাজনের আলোচনায় প্রবৃত্ত হতে হয়। মনে কর কয়েকজন ছাত্র পাঁচটি বিভিন্ন বিষয়ে পরীক্ষা দিয়ে তার ওপর নম্বর পেয়েছে। তাহলে, ঐ নম্বরগুলির ভিত্তিতে জানবার চেষ্টা করা যেতে পারে ঐ বিষয়গুলির মধ্যে কোন সংশ্রব আছে কিনা এবং তাদের মধ্যে দু-একটি বিষয় বেছে নিয়ে দেখা যেতে পারে তাদের ওপর অন্য চলগুলির প্রভাব দূর ক'রে দেওয়ার পরও তাদের মধ্যে লক্ষণীয় সহগতি আছে কিনা, এবং তার মাত্রা কী, ইত্যাদি। এ ছাড়া আরও একটি বিষয় দেখা যেতে পারে। সেটি হচ্ছে, একটি চলের ওপর অন্য চলগুলির একটি যৌথ বা সম্মিলিত প্রভাব আছে কি না এবং তার মাত্রাই বা কি। তাহলে, আমাদের অভিপ্রায় হবে এই প্রভাবটি সম্পর্কে যথেষ্ট জ্ঞান সঞ্চয় করা, যার সাহায্যে ঐ প্রভাবশীল চলগুলির প্রদত্ত মানের ভিত্তিতে প্রভাবিত চলটি সম্পর্কে কোন পূর্বাভাষ বা অনুমিতির চেষ্টা করা। এই সমস্যাগুলি এখন আমরা সংক্ষেপে আলোচনা ক'রব। তার আগে এটা ব'লে নেওয়া দরকার যে, প্রথমে এই চলগুলি সম্পর্কে লব্ধ রাশিতথ্যকে সংক্ষিপ্ত ও সার্থকভাবে প্রকাশ করার প্রচলিত ব্যবস্থাদি নিতে হবে, যেমন একচল ও দ্বিচল তথ্যসম্পর্কে নেওয়ার কথা আগে বলা হয়েছে।

যদি $p$ সংখ্যক চল $X_1, \ldots, X_i, \ldots X_p$ থাকে এবং $n$টি ব্যষ্টির $\alpha$-তম ব্যষ্টির জন্যে $X_i$-এর মান হয় $X_{i\alpha}$ $(i=1, \ldots, p\ ;\ \alpha=1, \ldots, n)$, তবে এই তথ্যকে $n$-টি সারি ও স্তম্ভযুক্ত একটি ম্যাট্রিক্সের সাহায্যে প্রকাশ করা যেতে পারে, যেটি হবে নিম্নরূপ :

$$\mathbf{X}=\begin{bmatrix} x_{11} & x_{21} & \cdots & x_{i1} & \cdots & x_{p1} \\ x_{12} & x_{22} & \cdots & x_{i2} & \cdots & x_{p2} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{1\alpha} & x_{\alpha 2} & \cdots & x_{i\alpha} & \cdots & x_{p\alpha} \\ \vdots & \vdots & & \vdots & & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{in} & \cdots & x_{pn} \end{bmatrix}.$$

যদি $n$ খুব বড় হয় তাহলে এই রাশিতথ্যকে পরিসংখ্যা বিভাজনের সাহায্যে প্রকাশ করা যেতে পারে। মনে কর সুবিধেমতো ভাবে মান বেছে বিভিন্ন চলের জন্যে শ্রেণী অন্তরসমূহ স্থির ক'রে $X_1, \ldots, X_i, \ldots, X_p$ চলগুলির জন্যে যথাক্রমে $k_1, k_2, \ldots k_i, \ldots, k_p$ টি শ্রেণী নির্ধারিত হ'ল। তাহলে মোট $k_1 \times \ldots \times k_i \times \ldots \times k_p$টি প্রকোষ্ঠে মোট পরিসংখ্যা $n$-কে বিভক্ত ক'রে ছড়িয়ে দেখাতে হবে এবং এই বিন্যাসটিই রচনা করবে $p$-সংখ্যক চলের যৌথ বা সম্মিলিত বিভাজন। এর থেকে প্রান্তীয় ও সর্তাধীন বিভাজনও নির্দেশ করা যায়। এ ছাড়া এ জাতীয় রাশিতথ্যকে চিত্র সাহায্যে প্রদর্শনেরও ব্যবস্থা রয়েছে। কিন্তু সংশ্লিষ্ট জটিলতার কথা মনে করে আমরা সে চেষ্টা ক'রব না। বর্তমান বিষয়ের প্রধানতঃ তুটি দিক নিয়ে আমরা পরপর আলোচনা ক'রব।

### 11.2 বহুল নির্ভরণ (Multiple Regression) :

মনে কর কয়েকটি ক্ষেতে বিভিন্ন পরিমাণে নাইট্রোজেন সার ($X_2$), ফস্‌ফেট সার ($X_3$) এবং বিভিন্ন পরিমাণে গোময় ($X_4$) প্রয়োগ ক'রে বিভিন্ন পরিমাণ ফসল ($X_1$) উৎপন্ন হ'ল। তাহলে সাধারণ বুদ্ধিতে বোঝা যায় যে, $X_2$, $X_3$ ও $X_4$-এর মানের পরিবর্তনের সঙ্গে সঙ্গে $X_1$-এর মানের হেরফের ঘটে এবং $X_1$ চলটি কোন না কোনভাবে অন্য চলগুলির ওপর নির্ভরশীল। এই সত্য অনুমান ক'রে একে কাজে লাগানোর চেষ্টা করা যেতে পারে। $X_2, X_3$, $X_4$-এর মান জানা থাকলে তাদের সাহায্যে $X_1$ সম্পর্কে পূর্বাভাষ করার চেষ্টা করা যেতে পারে এবং তার সাহায্যে জানা যেতে পারে $X_2$, $X_3$, $X_4$-এর কোন মিলিতমানের জন্যে $X_1$-এর সবচেয়ে প্রকৃষ্ট মান প্রত্যাশা করা যেতে পারে। ওপরের উদাহরণে যে তিনরকমের সারের উল্লেখ করা হয়েছে সেগুলি অনেক সময় আমাদের নিয়ন্ত্রণাধীন থাকে অর্থাৎ এগুলি আমাদের হাতে থাকে। কিন্তু উৎপন্ন দ্রব্যের ($X_1$) পরিমাণ স্বভাবতঃ অজ্ঞাত থাকে। তাই আমাদের উদ্দেশ্য হবে সারগুলির কোন বিশেষ বিশেষ পরিমাণ প্রয়োগের ফলে উৎপন্ন ফসলের প্রত্যাশিত পরিমাণ সম্পর্কে পূর্বানুমানের চেষ্টা করা। এইটিই হচ্ছে বহুল নির্ভরণের সমস্যা। এখানে কয়েকটি চলের ওপর অপর একটি চলের নির্ভরশীলতার সম্পর্কে তথ্য ব্যবহার ক'রে তার সাহায্যে ঐ চলটি সম্পর্কে পূর্বানুমানের চেষ্টা করা হয় প্রথমোক্ত চলগুলি সম্পর্কে জ্ঞানের ভিত্তিতে। ঐ শেষোক্ত চলটিকে ধরা হয় নির্ভরী চল (dependent) এবং প্রথমোক্ত চলদের

প্রত্যেককেই বলা হয় অনধীন চল (independent). এখানে চলগুলির অনধীনতা বা পরস্পর নির্ভরশীলতা অবশ্য গাণিতিক অর্থে ( সম্ভাবনাতাত্ত্বিক অর্থে নয় )। এই আলোচনায় আমাদের অনুসৃত পদ্ধতি হবে স্বনির্ভর চল $X_2$, $X_3,\ldots, X_p$-এর ওপর নির্ভরী চল $X_1$-এর একটি নির্ভরতার সম্পর্ক স্থাপন করা এবং তাকে একটি গাণিতিক সূত্রে প্রকাশ ক'রে তার সাহায্যে $X_2$, $X_3,\ldots, X_p$-এর প্রদত্ত মানসমূদয়ের জন্যে $X_1$-এর মান সম্পর্কে অনুমান করা। সাধারণতঃ এক্ষেত্রে একটি ঋজুরৈখিক অপেক্ষকের সাহায্যে $X_2, X_3,\ldots, X_p$-এর প্রদত্ত মানের জন্যে $X_1$-এর অনুমিত মান নির্ণয়ের চেষ্টা করা হয়। অর্থাৎ

$$X_{1\cdot 23\ldots p}=a+b_2X_2+b_3X_3+\cdots+b_pX_p$$

—এই ঋজুরৈখিক অপেক্ষকটিকে $X_2, X_3,\ldots, X_p$-এর প্রদত্তমানের ভিত্তিতে $X_1$ সম্পর্কে অনুমিতির উদ্দেশ্যে একটি প্রাক্‌কলন সূত্র হিসেবে নেওয়া হবে। এই সূত্রটিতে অজ্ঞাত অঙ্কগুলি হচ্ছে $a_1, b_2, b_3,\ldots, b_p$. কাজেই এই অপেক্ষককে রাশিতথ্যের ভিত্তিতে সম্পূর্ণ জানতে হলে এদের প্রাক্‌কলক নির্ণয় করতে হবে। এই উদ্দেশ্যে পূর্ববর্ণিত [ দশম পরিচ্ছেদ দ্রষ্টব্য ] লঘিষ্ঠ বর্গনীতিই প্রয়োগ করা হবে। অর্থাৎ $X_1$ সম্পর্কে অনুমিত মান যদি $X_{1\cdot 23\ldots p}$ হয়, তাহলে যে কোন $a$-তম ব্যষ্টির জন্যে এই সূত্রানুযায়ী অনুমানের ভ্রান্তি হবে

$$X_{1a}-X_{1\cdot 23\ldots pa}=X_{1a}-(a+b_2X_{2a}+\cdots+b_pX_{pa}).$$

এখানে $X_{ia}$ $(i=1,\ldots, p)$ হচ্ছে $a$-তম ব্যষ্টির জন্যে $(a=1,\ldots, n)$ $X_i$ চলের মান এবং $X_{1\cdot 23\ldots pa}$ হচ্ছে $a$-তম ব্যষ্টির জন্যে $X_{1\cdot 23\ldots p}$ চলের মান। এখন, লঘিষ্ঠ বর্গনীতি প্রয়োগ করতে হলে $a, b_2,\ldots, b_p$-কে এমনভাবে বেছে নিতে হবে যেন এই ভ্রান্তিগুলির বর্গের সমষ্টি সবচেয়ে কম হয় অর্থাৎ

$s=\sum_{a=1}^{n}(X_{1a}-X_{1\cdot 23\ldots pa})^2$-এর মান সর্বনিম্ন হয়। এ উদ্দেশ্যে অন্তর্কলন পদ্ধতি প্রয়োগ ক'রে $a, b_2,\ldots, b_p$-এর প্রাক্‌কলক পেতে হলে

$$\frac{\partial s}{\partial a}=0 \text{ অর্থাৎ } -2\sum(X_{1a}-a-b_2X_{2a}-\cdots-b_pX_{pa})=0$$

$$\frac{\partial s}{\partial b_2}=0 \text{ অর্থাৎ } -2\sum X_{2a}(X_{1a}-a-b_2X_{2a}-\cdots-b_pX_{pa})=0$$

$$\vdots$$

$$\frac{\partial s}{\partial b_p}=0 \text{ অর্থাৎ } -2\sum X_{pa}(X_{1a}-a-b_2X_{2a}-\cdots-b_pX_{pa})=0$$

—এই $p$-সংখ্যক নর্ম্যাল সমীকরণকে সমাধান করতে হয়। এখন ওগুলি দাঁড়ায়

$$\sum X_{1\alpha} = na + b_2 \sum X_{2\alpha} + \cdots + b_p \sum X_{p\alpha}$$

$$\sum X_{1\alpha} X_{2\alpha} = a \sum X_{2\alpha} + b_2 \sum X_{2\alpha}^2 + \cdots + b_p \sum X_{p\alpha} X_{2\alpha}$$

$$\vdots$$

$$\sum X_{1\alpha} X_{p\alpha} = a \sum X_{p\alpha} + b_2 \sum X_{2\alpha} X_{p\alpha} + \cdots + b_p \sum X_{p\alpha}^2$$

এখন যদি লেখা যায় $\frac{1}{n} \sum_{\alpha=1}^{n} x_{i\alpha} = \bar{x}_i$

এবং $s_{ij} = \frac{1}{n} \sum_{\alpha=1}^{n} (x_{i\alpha} - \bar{x}_i)(x_{i\alpha} - \bar{x}_j)$

তাহলে এগুলি দাঁড়াবে

$$\bar{x}_1 = a + b_2 \bar{x}_2 + \cdots + b_p \bar{x}_p$$

অর্থাৎ $a = \bar{x}_1 - b_2 \bar{x}_2 - \cdots - b_p \bar{x}_p,$ $\cdots$ (11.1)

এবং
$$\left.\begin{aligned} s_{12} &= b_2\, s_{22} + b_3\, s_{32} + \cdots + b_p\, s_{p2} \\ s_{13} &= b_2\, s_{23} + b_3\, s_{33} + \cdots + b_p\, s_{p3} \\ &\vdots \\ s_{1p} &= b_2\, s_{2p} + b_3\, s_{3p} + \cdots + b_p\, s_{pp} \end{aligned}\right\} \quad \cdots (11.2)$$

এই (11.2) সমীকরণগুলিকে ম্যাট্রিক্সের আকারে লেখা যায়

$$\begin{bmatrix} s_{12} \\ s_{13} \\ \cdot \\ \cdot \\ s_{1p} \end{bmatrix} = \begin{bmatrix} s_{22} & s_{32} \cdots s_{p2} \\ s_{23} & s_{33} \cdots s_{p3} \\ \cdot & \cdot \quad\quad \cdot \\ \cdot & \cdot \quad\quad \cdot \\ s_{2p} & s_{3p} \cdots s_{pp} \end{bmatrix} \begin{bmatrix} b_2 \\ b_3 \\ \cdot \\ \cdot \\ b_p \end{bmatrix}$$

অর্থাৎ

$$\begin{bmatrix} b_2 \\ b_3 \\ \cdot \\ \cdot \\ b_p \end{bmatrix} = \begin{bmatrix} s_{22} & s_{32} \cdots s_{p2} \\ s_{23} & s_{33} \cdots s_{p3} \\ \cdot & \\ \cdot & \\ s_{2p} & s_{3p} \cdots s_{pp} \end{bmatrix}^{-1} \begin{bmatrix} s_{12} \\ s_{13} \\ \cdot \\ \cdot \\ s_{1p} \end{bmatrix} \quad \cdots (11.3)$$

এখন, $|S| = \begin{vmatrix} s_{22} & s_{23} \cdots s_{2p} \\ s_{32} & s_{33} \cdots s_{3p} \\ \cdot & \\ \cdot & \\ s_{p2} & s_{p3} \cdots s_{pp} \end{vmatrix}$ এবং

$$|S|_j = \begin{vmatrix} s_{22} & s_{23} \cdots s_{2\overline{j-1}} & s_{21} & s_{2\overline{j+1}} \cdots s_{2p} \\ s_{32} & s_{33} \cdots s_{3\overline{j-1}} & s_{31} & s_{3\overline{j+1}} \cdots s_{3p} \\ \cdot & & & \\ \cdot & & & \\ s_{p2} & s_{p3} \cdots s_{p\overline{j-1}} & s_{p1} & s_{p\overline{j+1}} \cdots s_{pp} \end{vmatrix}$$

লিখে (11.3)-এর সমাধান ক'রে পাওয়া যায়

$$b_j = \frac{|S|_j}{|S|}. \qquad \cdots (11.4)$$

আবার, $x_i$ ও $x_j$-এর সহগাঙ্ককে $r_{ij}$ এবং $x_i$-এর ভেদমানকে $s_i^2 = s_{ii} = \frac{1}{n} \sum (x_{i\alpha} - \bar{x}_i)^2$ লিখে দেখা যায় যে আমরা লিখতে পারি

$x_i$ ও $x_j$-এর সহভেদমান $= \text{cov}\,(x_i, x_j) = s_{ij} = r_{ij}\, s_i\, s_j.$ ফলে আমরা (11.4) থেকে লিখতে পারি

$$b_j = (-1)^{j-2} \frac{s_1}{s_j} \begin{vmatrix} r_{21} & r_{22} \cdots r_{2\overline{j-1}} & r_{2\overline{j+1}} \cdots r_{2p} \\ r_{31} & r_{32} \cdots r_{3\overline{j-1}} & r_{3\overline{j+1}} \cdots r_{3p} \\ \cdot & & \\ \cdot & & \\ r_{p1} & r_{p2} \cdots r_{p\overline{j-1}} & r_{p\overline{j+1}} \cdots r_{pp} \end{vmatrix} \div \begin{vmatrix} r_{22} & r_{23} \cdots r_{2p} \\ r_{32} & r_{33} \cdots r_{3p} \\ \cdot & \\ \cdot & \\ r_{p2} & r_{p3} \cdots r_{pp} \end{vmatrix} \qquad \cdots (11.5)$$

এখন, $(R) = \begin{bmatrix} r_{11} & r_{12} \cdots r_{1p} \\ r_{21} & r_{22} \cdots r_{2p} \\ \cdot & \\ \cdot & \\ r_{p1} & r_{p2} \cdots r_{pp} \end{bmatrix}$ কে সহগাঙ্ক ম্যাট্রিক্স, এবং

$$R = \begin{vmatrix} r_{11} & r_{12} \cdots r_{1p} \\ r_{21} & r_{22} \cdots r_{2p} \\ \cdot & \\ \cdot & \\ r_{p1} & r_{p2} \cdots r_{pp} \end{vmatrix}$$

-কে সহগাঙ্ক ডিটারমিন্যান্ট ব'লে উল্লেখ ক'রব এবং $R$-এ $r_{ij}$-এর সহ-উৎপাদক (co-factor)-কে $R_{ij}$ দ্বারা চিহ্নিত ক'রব। তাহলে লেখা যাবে

$$b_j = (-1)^{j-2} \times (-1)^{j+1} \frac{S_1}{S_j} . \frac{R_{1j}}{R_{11}}$$

অর্থাৎ $$b_j = (-1)^{2j-1} \frac{S_1}{S_j} . \frac{R_{1j}}{R_{11}} = -\frac{S_1}{S_j} . \frac{R_{1j}}{R_{11}} \ (j = 2, 3, \ldots, p). \quad \cdots (11.6)$$

তাহলে, $$a = \overline{X}_1 - \sum_{j=2}^{p} b_j \overline{X}_j = \overline{X}_1 + \sum_{j=2}^{p} \frac{R_{1j}}{R_{11}} . \frac{S_1}{S_j} \overline{X}_j. \quad \cdots (11.7)$$

ফলে, শেষপর্যন্ত অনুমিতিসূত্রটি দাঁড়ায়

$$\hat{X}_{1 \cdot 23 \ldots p} = \overline{X}_1 + b_2(X_1 - \overline{X}_2) + \cdots + b_p(X_p - \overline{X}_p)$$

বা $$\hat{X}_{1 \cdot 23 \ldots p} = \overline{X}_1 - \frac{R_{12}}{R_{11}} . \frac{S_1}{S_2} (X_2 - \overline{X}_1) - \cdots - \frac{R_{1p}}{R_{11}} \frac{S_1}{S_p} (X_p - \overline{X}_p).$$

এই সমীকরণটিকে বলা হয় বহুল-নির্ভরণ-সমীকরণ (Multiple Regression Equation). এর সাহায্যে $X_2, X_3, \ldots, X_p$ এই কটি স্বনির্ভর চলের ওপর নির্ভরশীল অপর একটি চল $X_1$-এর ঋজুরৈখিক নির্ভরতা লঘিষ্ঠ বর্গনীতি অনুযায়ী নির্ধারিত ও প্রকাশিত হয়। একে $X_2, X_3, \ldots X_p$-এর ওপর $X_1$-এর বহুল নির্ভরণী ঋজুরেখাও (Multiple Regression line) বলা হয়। এর সাহায্যে $X_2, X_3, \ldots, X_p$ সম্পর্কে যে কোন প্রদত্ত মান অনুযায়ী প্রাপ্ত মানকে $X_1$-এর অনুমাপক হিসেবে ব্যবহার করা হয়। এখানে $b_j = -\frac{R_{1j}}{R_{11}} . \frac{S_1}{S_j} \ (j = 2, 3, \ldots, p)$ হচ্ছে $X_j$-এর ওপর $X_1$-এর আংশিক নির্ভরণাঙ্ক (Partial Regression Coefficient). এরূপ নাম দেওয়ার কারণ $X_j$ ছাড়া আরও চল রয়েছে এবং তাদের প্রত্যেকেরই ওপরে $X_1$-এর নির্ভরতা রয়েছে। অনেক সময় লেখা হয় $b_j = b_{1j \cdot 23 \ldots \overline{j-1}\, \overline{j+2} \cdots p}$. কারণ, তাতে বোঝা যায় কোন্ চলের ওপর $X_1$-এর নির্ভরণ বিবেচনা করা হচ্ছে এবং আনুষঙ্গিক অন্যান্য চলগুলিই বা কী কী। এই $b_{1j \cdot 23 \ldots \overline{j-1}\, \overline{j+2} \ldots p}$ নির্দেশ করে $X_j$-এর প্রতি একক (unit) পরিবর্তনে $X_{1 \cdot 23 \ldots p}$ কতটুকু পরিবর্তিত হবে, অবশ্য যদি সেই সঙ্গে অন্য চলগুলির মান স্থির থাকে।

## 11.3 বহুল সহগতি (Multiple Correlation):

বহুল নির্ভরণতত্ত্বের আলোচনায় আমরা দেখবার চেষ্টা করেছি কিভাবে কয়েকটি স্বনির্ভর চলের ওপর অপর একটি চলের নির্ভরশীলতা লক্ষ্য ক'রে প্রথমোক্ত চলদের সম্পর্কে জ্ঞাত তথ্য ব্যবহার ক'রে শেষোক্ত চলটি সম্পর্কে অনুমান করা সম্ভব। এখন, এ জাতীয় রাশিতথ্যকে আরও একটি বিষয়ের চর্চায় নিয়োজিত করা যেতে পারে। আমরা দেখবার চেষ্টা করতে পারি প্রথমোক্ত চলগুলি যৌথভাবে শেষোক্ত চলটিকে কেমন করে এবং কতখানি প্রভাবিত করতে পারে। অর্থাৎ আমরা মাপবার চেষ্টা করতে পারি কিভাবে এবং কতখানি সার্থকতা এবং গভীরতার সঙ্গে কয়েকটি পরস্পর স্বনির্ভর চল অপর একটি চলের ওপর তাদের পৃথক্ পৃথক্ প্রভাব একযোগে বিস্তার করতে পারে। এই উদ্দেশ্যে বহুল সহগতির (Multiple Correlation) ব্যবহার হয়ে থাকে।

আমরা আগে দেখেছি যে, যদি দুটি মাত্র চল থাকে তাহ'লে তারা পরস্পর কতখানি সংশ্রবযুক্ত তা তাদের সহগাঙ্ক $r$-এর মান থেকে জানা যায়। কিন্তু নির্ভরণতত্ত্বের আলোচনাসূত্রে আরও দেখা গেছে যে, বাস্তবিক, $|r|$-এর মানকে বিবেচনা করেই আমরা জানতে পারি যদি দুটি চলের একটিকে অপরটির ওপর নির্ভরশীল ব'লে মনে করা হয় এবং তাদের সম্পর্ক যদি অন্ততঃ মোটামুটিভাবে ঋজুরৈখিক প্রকৃতিবিশিষ্ট ব'লে ধরা যায়, তাহলে স্বনির্ভর চলটি কতটুকু সার্থক এবং তীব্রভাবে অপর চলটির ওপর তার প্রভাব বিস্তার করতে পারে। আমরা আরও দেখেছি যে, $Y$ যদি নির্ভরশীল ও $X$ যদি স্বনির্ভর চল হয় এবং $\hat{Y}=A+BX$ সমীকরণ সম্বলিত ঋজুরেখাটির সাহায্যে $Y$-এর ওপর $X$-এর প্রভাব রয়েছে এই ধারণায় $Y$ সম্পর্কে $X$-এর সাহায্যে যদি অনুমান করতে যাই, তাহলে পাওয়া যায় $|r|=r_{Y\hat{Y}}$. এর থেকে বোঝা যাচ্ছে যে, $|r|$-এর মান থেকেই আমরা জানতে পারি $X$ কতখানি নিবিড়ভাবে $Y$-এর ওপর তার প্রভাব বিস্তার করে অবশ্য যদি $X$ ও $Y$-এর সম্পর্ক অন্ততঃ আসন্নভাবেও ঋজুরৈখিক ধরনের হয়। এই সমস্ত বিষয়গুলি মনে রেখেই $X_1$ এবং $X_{1.23\ldots p}$ এই দুটি চলের সহগাঙ্কটির সাহায্যে পরিমাপ করার চেষ্টা করা হয় $X_2, X_3, \ldots, X_p$ চলগুলি একযোগে $X_1$ চলটির ওপর কতখানি সার্থক ও গভীরভাবে তাদের সম্মিলিত প্রভাব বিস্তার করে। অবশ্য, এটা ধরে নেওয়া হয় যে, $X_2, X_3, \ldots, X_p$-এর ওপর $X_1$-এর নির্ভরতার প্রকৃতি অন্ততঃ আসন্নভাবেও ঋজুরৈখিক ধরনের। এই সহগাঙ্ককে

বলা হয় $X_2$, $X_3, \ldots, X_p$-এর ওপর $X_1$-এর বহুল সহগাঙ্ক। তাহলে $X_2, X_3, \ldots, X_p$-এর ওপর $X_1$-এর বহুল সহগাঙ্ক বলতে আমরা বুঝব লঘিষ্ঠ বর্গনীতি অনুযায়ী নির্ণীত $X_2, \ldots, X_p$-এর ওপর $X_1$-এর সাযুজ্য রক্ষাকারী ঋজুরৈখিক নির্ভরণ অপেক্ষকের (fitted linear regression function) সঙ্গে $X_1$-এর সহগাঙ্ক। একে $r_{1\cdot 23\ldots p}$ সংকেতসূত্রে প্রকাশ করা হবে এবং এর সূত্র হচ্ছে

$$r_{1\cdot 23\ldots p}=\frac{\text{cov}\,(X_1,\, X_{1\cdot 23\ldots p})}{\sqrt{V(X_1)}\cdot\sqrt{V(X_{1\cdot 23\ldots p})}}. \qquad \cdots (11.9)$$

এখানে $X_{1\cdot 23\ldots p}=a+b_2X_2+\cdots+b_pX_p$ এবং $a, b_2, \ldots, b_p$-এর প্রাক্কলক লঘিষ্ঠ বর্গনীতি অনুযায়ী নির্ধারিত ( দ্রষ্টব্য : 11.2 অনুচ্ছেদ )। আমরা দেখেছি ( অনুচ্ছেদ 11.2) যে, $V(X_1)=S_1^2$ এবং $X_{1\cdot 23\ldots p}$-এর পূর্ণকাঙ্কচয় $a, b_2, \ldots b_p$-এর প্রাক্কলক নির্ণয়ে ব্যবহৃত নর্ম্যাল সমীকরণগুলি হচ্ছে :

$$\sum_a x_{1\cdot 23\ldots p}=0,\ \sum_a X_2\, x_{1\cdot 23\ldots p}=0, \ldots, \sum_a X_p x_{1\cdot 23\ldots p}=0.$$

এখানে $x_{1\cdot 23\ldots p}=X_1-X_{1\cdot 23\ldots p}$ লেখা হয়েছে। কাজেই

$$0=\frac{1}{n}\sum_a x_{1\cdot 23\ldots pa}=\frac{1}{n}\sum_a(X_{1a}-X_{1\cdot 23\ldots pa})$$

$$=\bar{X}_1-\frac{1}{n}\sum_a X_{1\cdot 23\ldots p}.$$

ফলে, $$\bar{X}_{1\cdot 23\ldots p}=\frac{1}{n}\sum X_{1\cdot 23\ldots pa}=\bar{X}_1 \qquad \cdots (11.10)$$

আবার $X_1=X_{1\cdot 23\ldots p}+x_{1\cdot 23\ldots p}$ থেকে পাওয়া যায়

$$\text{Cov}\,(X_1, X_{1\cdot 23\ldots p})=\text{cov}\,(X_{1\cdot 23\ldots p}+x_{1\cdot 23\ldots p},\, X_{1\cdot 23\ldots p})$$
$$=V(X_{1\cdot 23\ldots p})+\text{cov}\,(X_{1\cdot 23\ldots p},\, x_{1\cdot 23\ldots p})$$

[ অনুশীলনী 11.6 দ্রষ্টব্য ]

আবার, $\text{cov}\,(x_{1\cdot 23\ldots p},\, X_{1\cdot 23\ldots p})$

$$=\frac{1}{n}\sum(X_{1\cdot 23\ldots pa}-\bar{X}_1)\,x_{1\cdot 23\ldots pa}$$

$$=\frac{1}{n}\sum X_{1\cdot 23\ldots pa}x_{1\cdot 23\ldots pa}-\bar{X}_1\cdot\frac{1}{n}\sum x_{1\cdot 23\ldots pa}$$

$$=\frac{1}{n}\sum(a+b_2 X_{2\alpha}+\cdots+b_p X_{p\alpha})$$

$$x_{1\cdot 23\cdots p\alpha}-\overline{X}_1\cdot\frac{1}{n}\sum x_{1\cdot 23\cdots p\alpha}$$

$$=\frac{1}{n}\Big[a\sum x_{1\cdot 23\cdots p\alpha}+b_2\sum X_{2\alpha}.x_{1\cdot 23\cdots p\alpha}$$

$$+\cdots+b_p\sum X_{p\alpha}x_{1\cdot 23\cdots p\alpha}\Big]$$

$=0$ [ নর্ম্যাল সমীকরণ ব্যবহার ক'রে ].

সুতরাং $\text{cov}\,(X_1, X_{1\cdot 23\cdots p})=V(X_{1\cdot 23\cdots p})$.

কাজেই $r_{1\cdot 23\cdots p}=\sqrt{\dfrac{V(X_{1\cdot 23\cdots p})}{V(X_1)}}=\dfrac{1}{s_1}\cdot\sqrt{V(X_{1\cdot 23\cdots p})}$.

এখন, $V(X_{1\cdot 23\cdots p})=\text{cov}\,(X_1, X_{1\cdot 23\cdots p})$

$$=\frac{1}{n}\sum_\alpha(X_{1\alpha}-\overline{X}_1)(X_{1\cdot 23\cdots p\alpha}-\overline{X}_1)$$

$$=\frac{1}{n}\sum_\alpha(X_{1\alpha}-\overline{X}_1)\Big[\Big\{\overline{X}_1-\frac{S_1}{S_2}\frac{R_{12}}{R_{11}}(X_{2\alpha}-\overline{X}_2)$$

$$-\frac{S_1}{S_3}\frac{R_{13}}{R_{11}}(X_{3\alpha}-\overline{X}_3)$$

$$\cdots-\frac{S_1}{S_p}\frac{R_{1p}}{R_{11}}(X_{p\alpha}-\overline{X}_p)\Big\}-\overline{X}_1\Big]$$

[ (11.8) দ্রষ্টব্য ]

$$=-\frac{S_1}{S_2}.\frac{R_{12}}{R_{11}}\frac{1}{n}\sum_\alpha(X_{1\alpha}-\overline{X}_1)(X_{2\alpha}-\overline{X}_2)$$

$$-\frac{S_1}{S_p}.\frac{R_{13}}{R_{11}}.\frac{1}{n}\sum_\alpha(X_{1\alpha}-\overline{X}_1)(X_{3\alpha}-\overline{X}_3)$$

$$\cdots-\frac{S_1}{S_p.}\frac{R_{1p}}{R_{11}}\frac{1}{n}\sum_\alpha(X_{1\alpha}-\overline{X}_1)(X_{p\alpha}-\overline{X}_p)$$

$$=-\frac{S_1}{S_2}.\frac{R_{12}}{R_{11}}S_{12}-\frac{S_1}{S_3}.\frac{R_{13}}{R_{11}}S_{13}$$

$$-\cdots-\frac{S_1}{S_p}\frac{R_{1p}}{R_{11}}S_{1p}$$

$$= -\frac{S_1^2}{R_{11}}(r_{12}R_{12}+r_{13}R_{13}+\cdots+r_{1p}R_{1p})$$

$$= -\frac{S_1^2}{R_{11}}(r_{11}R_{11}+r_{12}R_{12}$$

$$+\cdots+r_{1p}R_{1p}-r_{11}R_{11})$$

$$= -\frac{S_1^2}{R_{11}}\left(R-r_{11}R_{11}\right)=S_1^2\left(1-\frac{R}{R_{11}}\right).$$

তাই $r_{1\cdot 23\ldots p}=\frac{1}{S_1}\sqrt{V(X_{1\cdot 23\ldots p})}$

$$=\sqrt{1-\frac{R}{R_{11}}} \quad \cdots \quad \cdots \quad (11.11)$$

তাই $r^2_{1\cdot 23\ldots p}=\left(1-\frac{R}{R_{11}}\right).$ ··· ··· (11.12)

এখন, $V(X_1)=V(X_{1\cdot 23\ldots p})+V(x_{1\cdot 23\ldots p})$

$$+2\,\text{cov}\,(X_{1\cdot 23\text{-}p}, x_{1\cdot 23\ldots p})$$

$$=V(X_{1\cdot 23\ldots p})+V(x_{1\cdot 23\ldots p}).$$

ফলে, $V(x_{1\cdot 23\ldots p})=V(X_1)-V(X_{1\cdot 23\ldots p})$

$$=S_1^2-S_1^2r_{1\cdot 23\ldots p}=S_1^2(1-r^2_{1\cdot 23\ldots p})$$

$$=S^2\left[1-\left(1-\frac{R}{R_{11}}\right)\right]=S_1^2\cdot\frac{R}{R_{11}}.$$

আবার, $V(x_{1\cdot 23\ldots p})$-কে $s^2_{1\cdot 23\ldots p}$ সংকেতসূত্রে প্রকাশ করলে লেখা যায়,

$$r^2_{1\cdot 23\ldots p}=1-\frac{s^2_{1\cdot 23\cdots p}}{s_1^2}=1-\frac{V(x_{1\cdot 23\ldots p})}{V(X_1)}.$$

এখন, বলা যেতে পারে যে, $X_{1\cdot 23\ldots p}$ নির্দেশ করছে $X_1$ চলের মানের যে অংশ $X_2$, $X_3$,...., $X_p$-এর উপর ঋজুরৈখিক নির্ভরণের মাধ্যমে নির্ণীত হয়েছে এবং $x_{1\cdot 23\ldots p}$ হচ্ছে তৎপরবর্তী উদ্বৃত্তাংশ (Residual). তাহলে, $r^2_{1\cdot 23\ldots p}=\frac{V(X_{1\cdot 23\ldots p})}{V(X_1)}$ হচ্ছে $X_1$-এর সমগ্র প্রভেদের যে ভগ্নাংশ $X_2$, $X_3$, ..., $X_p$-এর ওপর $X_1$-এর ঋজুরৈখিক নির্ভরণসূত্রের সাহায্যে ব্যাখ্যাত হয়েছে এবং $1-r^2_{1\cdot 23\ldots p}=\frac{V(x_{1\cdot 23\ldots p})}{V(X_1)}$ হচ্ছে $X_1$-এর সমগ্র প্রভেদের সেই ভগ্নাংশ যা ওভাবে ব্যাখ্যাত হয়নি। কাজেই $r_{1\cdot 23\ldots p}$-কে ভাবা যেতে পারে এমন

একটি মাপনাঙ্ক বা মাপক হিসেবে যার সাহায্যে আন্দাজ করা যায় ঋজুরৈখিক নির্ভরণ অপেক্ষক $X_{1\cdot 23\ldots p}$, $X_1$ সম্পর্কে $X_2, X_3, \ldots, X_p$-এর সাহায্যে অনুমাপক হিসেবে কতটা সার্থক। $r_{1\cdot 23\ldots p}$-এর মান যত বাড়বে $s^2_{1\cdot 23\ldots p} = V(x_{1\cdot 23\ldots p})$-এর মান তত কমবে এবং $r_{1\cdot 23\ldots p}$ যখন সর্বোচ্চ মান ($=+1$) গ্রহণ করবে তখন $V(x_{1\cdot 23\ldots p})=0$ অর্থাৎ প্রত্যেক $a=1, 2, \ldots, n$-এর জন্যে $x_{1\cdot 23\ldots pa}=\bar{x}_{1\cdot 23\ldots p}=0$ or $X_{1a}=X_{1\cdot 23\ldots pa}$ হবে। ফলে, এক্ষেত্রে পূর্বাভাষ সূত্র (Forecasting formula) $X_{1\cdot 23\ldots p}$ থেকে প্রত্যেক $X_2, X_3, \ldots X_p$-এর জন্যে $X_1$-এর আসল মানটিই পাওয়া যাবে এবং সেক্ষেত্রে নিশ্চয়ই বলা যাবে যে, $X_{1\cdot 23\ldots p}$ অনুমাপক হিসেবে সবচেয়ে কার্যকরী। পক্ষান্তরে, $r_{1\cdot 23\ldots p}$ যত কমতে থাকবে, নির্দিষ্ট $V(X_1)$-এর মানের জন্যে $V(x_{1\cdot 23\ldots p})$-এর মান তত বাড়তে থাকবে এবং $V(X_{1\cdot 23\ldots p})$-এর মান তত কমতে থাকবে অর্থাৎ $X_{1\cdot 23\ldots pa}$-এর মান ততই $\bar{X}_{1\cdot 23\ldots p}=\bar{X}_1$-এর নিকটবর্তী হতে থাকবে এবং চরম সীমায় যখন $r_{1\cdot 23\ldots p}=0$ হবে তখন প্রত্যেক $a$-র জন্যে $X_{1\cdot 23\ldots pa}=\bar{X}_1$ হবে। এর অর্থ হবে এই যে, $X_2, X_3, \ldots, X_p$ সম্পর্কে কোন জ্ঞানই $X_1$ সম্পর্কে অনুমানে আমাদের কোন সাহায্য করবে না, অবশ্য যদি সে উদ্দেশ্যে আমরা ঋজুরৈখিক নির্ভরণসূত্র $X_{1\cdot 23\ldots p}$-কে অনুমান মাধ্যম হিসেবে ব্যবহার করি।

বহুল সহগাঙ্ক $r_{1\cdot 23\ldots p}$ সম্পর্কে একটি উল্লেখযোগ্য বিষয় হচ্ছে এই যে, এর মান 0 ও 1-এর মধ্যে সীমাবদ্ধ অর্থাৎ $-1$ থেকে 0-এর মধ্যে এর কোন মান থাকতে পারে না। এর কারণ এই যে, $r_{1\cdot 23\ldots p}$ হচ্ছে দুটি প্রমাণ-বিচ্যুতির অনুপাত।

## 11.4 আংশিক সহগতি (Partial Correlation) :

অনেক সময় এমন হয়ে থাকে যে, আমরা যে দুটি মুখ্য চল $X_1$ ও $X_2$ সম্পর্কে আলোচনা করি তাদের মধ্যে লক্ষিত সহগতি কোন কার্যকারণ সূত্রের অস্তিত্ব সূচিত করে না। বরঞ্চ অনেক সময়ই এমন দেখা যায় যে, অন্য কয়েকটি চল $X_3$, $X_4, \ldots$ ইত্যাদির অস্তিত্ব থাকে যারা $X_1$ ও $X_2$ উভয়ের সঙ্গে সংস্রবযুক্ত এবং সেজন্যেই ঐ সহগতির মাধ্যমেই $X_1$ ও $X_2$-এর মধ্যেও পারস্পরিক সহগতি পরিলক্ষিত হয়। স্বভাবতঃই এক্ষেত্রে আমাদের জানতে কৌতূহল হবে $X_3$, $X_4$ ইত্যাদি বহিঃস্থ চলগুলির প্রভাব যদি $X_1$ ও $X_2$ উভয়ের ওপর থেকেই বিদূরিত

করা হয়, তাহলেও $X_1$ ও $X_2$-এর মধ্যে কোন সহগতি অবশিষ্ট থাকবে কি না। এভাবে $X_1$ ও $X_2$-কে $X_3$, $X_4$ ইত্যাদি চলের প্রভাব থেকে মুক্ত ক'রে নিয়ে তাদের যে সহগতি নির্ণয় করা হয় তাকে বলে $X_1$ ও $X_2$-এর মধ্যে আংশিক সহগতি বা নীট (partial or net) সহগতি। আর, পূর্বে আলোচিত $X_1$ ও $X_2$-এর সহগতিকে ( যখন তাদের ওপর $X_3, X_4, \ldots, X_p$ চলগুলির প্রভাব সম্পর্কে আমরা উদাসীন থাকি ) তার বৈপরীত্যে মোট বা পূর্ণ সহগতি (total correlation) বলা যেতে পারে। এখন, $X_1$ ও $X_2$-এর ওপর $X_3$, $X_4, \ldots, X_p$-এর প্রভাব কীভাবে দূর করা যাবে ? অবশ্যই সম্পূর্ণ সার্থকভাবে তা করা যাবে না। যতটুকু করা যাবে তা হচ্ছে এই যে, আমরা পৃথক্ পৃথক্ ভাবে $X_3, X_4, \ldots, X_p$-এর ওপর $X_1$ এবং $X_2$-এর ঋজুরৈখিক বহুল নির্ভরণ সূত্র $X_{1\cdot 34\ldots p}$ ও $X_{2\cdot 34\ldots p}$ নির্ণয় করতে পারি এবং তাদের থেকে দুটি উদ্বৃত্তাংশ $X_1 - X_{1\cdot 34\ldots p} = x_{1\cdot 34\ldots p}$ ও $X_2 - X_{2\cdot 34\ldots p} = x_{2\cdot 34\ldots p}$ নির্ণয় করতে পারি। বলা বাহুল্য $X_{1\cdot 34\ldots p}$ ও $X_{2\cdot 34\ldots p}$ সূত্রদুটি লঘিষ্ঠ বর্গনীতি অনুযায়ীই নির্ধারিত হবে। এখন এই উদ্বৃত্তাংশ দুটির মধ্যে যে সহগতি আছে তাকেই $X_1$ ও $X_2$-এর মধ্যে আংশিক সহগতি (partial correlation) বলা হবে। এই আংশিক সহগতি মাপনে ব্যবহৃত অঙ্কটিকে $X_1$ ও $X_2$-এর আংশিক সহগাঙ্ক (partial correlation coefficient) বলা হয় এবং স্বভাবতঃই তার সংজ্ঞা হচ্ছে

$$r_{12\cdot 34\ldots p} = \frac{\text{cov}\,(x_{1\cdot 34\ldots p}, x_{2\cdot 34\ldots p})}{\sqrt{V(x_{1\cdot 34\ldots p})}\,\sqrt{V(x_{2\cdot 34\ldots p})}}.$$

লেখা যাক

$$X_{1\cdot 34\ldots p} = c + b_{13\cdot 4\ldots p}X_3 + b_{14\cdot 35\ldots p}X_4 + \cdots + b_{1p\cdot 34\ldots \overline{(p-1)}}X_p$$

এবং $$X_{2\cdot 34\ldots p} = d + b_{23\cdot 4\ldots x}X_3 + b_{24\cdot 35\ldots p4}X_4 + \cdots + b_{2p\cdot 34\ldots \overline{(p-1)}}X_p.$$

এগুলি হচ্ছে যথাক্রমে $X_3, X_4, \ldots X_p$-এর ওপর $X_1$ ও $X_2$-এর লঘিষ্ঠ বর্গনীতি অনুযায়ী নির্ধারিত ঋজুরৈখিক নির্ভরণ সূত্র। তাহলে, নর্ম্যাল সমীকরণগুলি দাঁড়াবে

$$\sum_a x_{1\cdot 34\ldots pa} = 0, \quad \sum_a X_3.x_{1\cdot 34\ldots p} = 0\ldots\ldots,$$

$$\sum_a X_{pa}.x_{1\cdot 34\ldots pa} = 0$$

এবং $\sum_{\alpha} x_{2\cdot 34\ldots p\alpha}=0,\ \sum_{\alpha} X_{3\alpha}.x_{2\cdot 34\ldots p\alpha}=0,\ldots\ldots,$

$$\sum_{\alpha} X_{p\alpha}.x_{2\cdot 34\ldots p\alpha}=0.$$

[ উল্লেখ করা যেতে পারে যে, $x_{1\cdot 34\ldots p}=X_1-X_{1\cdot 34\ldots p}$ এবং $x_{2\cdot 34\ldots p}=X_2-X_{2\cdot 34\ldots p}$ হচ্ছে যথাক্রমে $X_1$ ও $X_2$-এর সেই অংশ যা লঘিষ্ঠ বর্গনীতি অনুযায়ী নির্ণীত $X_3, X_4, \ldots, X_p$-এর ওপর তাদের বহুল নির্ভরণ-ঋজুরেখা দ্বারা অনুমিত অংশ তাদের থেকে বিচ্ছিন্ন করার পরবর্তী উদ্‌বৃত্তাংশ। ]

তাহলে, $\bar{x}_{1\cdot 34\ldots p}=\frac{1}{n}\sum x_{1\cdot 34\ldots p\,\alpha}=0$ ও

$\bar{x}_{2\cdot 34\ldots p}=\frac{1}{n}\sum x_{2\cdot 34\ldots p\,\alpha}=0$ এবং $i=3,\ 4,\ldots p$-এর জন্যে

$$\text{cov}\,(x_{1\cdot 34\ldots p},\ X_i)=\frac{1}{n}\sum_{\alpha}(X_{i\alpha}-\bar{X}_i)\,x_{1\cdot 34\ldots p\alpha}$$

$$=\frac{1}{n}\sum_{\alpha} X_{i\alpha}\,x_{1\cdot 34\ldots p\alpha}-\bar{X}_i\,\frac{1}{n}\sum x_{1\cdot 34\ldots p\alpha}=0,$$ এবং, তদ্রূপ

$\text{cov}\,(x_{2\cdot 34\ldots p}, X_i)=0$ অর্থাৎ $x_{1\cdot 34\ldots p}$ ও $x_{2\cdot 34\ldots p}$ উভয়েই

$X_3, X_4, \ldots, X_p$-এর সঙ্গে সহগতিমুক্ত। এর থেকে আমরা ধরে নিতে পারি যে, $x_{1\cdot 34\ldots p}$ ও $x_{2\cdot 34\ldots p}$ হচ্ছে মোটামুটিভাবে $X_1$ ও $X_2$-এর সেই অংশ যা $X_3, X_4,\ldots X_p$-এর প্রভাব থেকে মুক্ত। এইজন্যেই $x_{1\cdot 34\ldots p}$ ও $x_{2\cdot 34\ldots p}$-এর সহগাঙ্ককে $X_1$ ও $X_2$-এর আংশিক সহগাঙ্ক ( তাদের ওপর $X_3$, $X_4, \ldots, X_p$-এর প্রভাব বিদূরিত করার পর ) হিসেবে মেনে নেওয়া যায়।

$X_3, \ldots, X_p$-এর ওপর $X_1$-এর নির্ভরণসূত্র বিবেচনা করলে প্রাসঙ্গিক সহগতি-ডিটারমিন্যাণ্ট হচ্ছে

$$R^{(2)}=\begin{vmatrix} r_{11} & r_{13}\ldots r_{1p} \\ r_{31} & r_{33}\ldots r_{3p} \\ \cdot & \cdot\quad\;\; \cdot \\ \cdot & \cdot\quad\;\; \cdot \\ r_{p1} & r_{p3}\ldots r_{pp} \end{vmatrix}.$$

লক্ষণীয় যে, $R^{(2)}$ হচ্ছে $R = \begin{vmatrix} r_{11} & r_{12} & r_{13} \ldots r_{1p} \\ r_{21} & r_{22} & r_{23} \ldots r_{2p} \\ r_{31} & r_{32} & r_{33} \ldots r_{3p} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ r_{p1} & r_{p2} & r_{p3} \ldots r_{pp} \end{vmatrix}$

থেকে দ্বিতীয় সারি ও দ্বিতীয় স্তম্ভ অপসারিত করার পর অবশিষ্টাংশ। কাজেই $V(x_{1\cdot 34\ldots p}) = S_1^2 \dfrac{R^{(2)}}{R_{11}^{(2)}}$. উল্লেখ্য যে, $R^{(2)}$ হচ্ছে $R_{22}$ অর্থাৎ $R$-এ $r_{11}$-এর সহ-উৎপাদক (co-factor) ও $R_{11}^{(2)}$ হচ্ছে $R^{(2)}$-তে $r_{11}$-এর সহ-উৎপাদক। তেমনি, যখন $X_3, X_4, \ldots X_p$-এর ওপর $X_2$-এর নির্ভরণসূত্র নির্ণয় করা হয়, তখন মোট যে সহগতি-ডিটারমিন্যাণ্ট ব্যবহার হয় তা হচ্ছে

$R^{(1)} = \begin{vmatrix} r_{22} & r_{23} \ldots r_{2p} \\ r_{32} & r_{33} \ldots r_{3p} \\ \cdot & \\ \cdot & \\ r_{p2} & r_{p3} \ldots r_{pp} \end{vmatrix}$. কাজেই $V(x_{2\cdot 34\ldots p}) = S_2^2 \dfrac{R^{(1)}}{R_{22}^{(1)}}$.

তাহলে, $\text{cov}\,(x_{1\cdot 34\ldots p}, x_{2\cdot 34\ldots p}) = \dfrac{1}{n} \sum_{\alpha} x_{1\cdot 34\ldots p\alpha}\, x_{2\cdot 34\ldots p\alpha}$

$$= \frac{1}{n} \sum_{\alpha} x_{2\cdot 34\ldots p\alpha} \Big[ X_{1\alpha} - c - b_{13\cdot 4\ldots p} X_{3\alpha} - \cdots$$

$$\cdots - b_{1p\cdot 34\ldots \overline{p-1}}\, X_{p\alpha} \Big]$$

$$= \frac{1}{n} \sum_{\alpha} X_{1\alpha}\, x_{2\cdot 34\ldots p\alpha} = \frac{1}{n} \sum (X_1 - \bar{X}_1)(x_{2\cdot 34\ldots p\alpha})$$

[ নর্ম্যাল সমীকরণ ব্যবহার ক'রে ]

$$= \frac{1}{n} \sum_{\alpha} (X_{1\alpha} - \bar{X}_1)(X_{2\alpha} - X_{2\cdot 34\ldots p\alpha})$$

$$= \frac{1}{n} \sum_{\alpha} (X_{1\alpha} - \bar{X}_1)[X_{2\alpha} - \{\bar{X}_2 - \frac{S_2}{S_3} \frac{R_{23}^{(1)}}{R_{22}^{(1)}} (X_{3\alpha} - \bar{X}_3)$$

$$- \cdots - \frac{S_2}{S_p} \frac{R_{2p}^{(1)}}{R_{22}^{(1)}} (X_{p\alpha} - \bar{X}_p)\}]$$

$$= \frac{1}{n}\sum_{a}(X_{1a}-\bar{X}_1)(X_{2a}-\bar{X}_2)$$

$$+\frac{S_2}{S_3}\cdot\frac{R_{23}^{(1)}}{R_{22}^{(1)}}\frac{1}{n}\sum_{a}(X_{1a}-\bar{X}_1)(X_{3a}-\bar{X}_p)$$

$$+\cdots+\frac{S_2}{S_p}\cdot\frac{R_{2p}^{(1)}}{R_{22}^{(1)}}\frac{1}{n}\sum_{a}(X_{1a}-\bar{X}_1)(X_{pa}-\bar{X}_p)$$

$$= r_{12}\,S_1S_2+\frac{S_2}{S_3}\cdot\frac{R_{23}^{(1)}}{R_{22}^{(1)}}r_{13}\,S_1S_3+\frac{S_2}{S_4}\frac{R_{24}^{(1)}}{R_{22}^{(1)}}r_{14}\,S_1S_4$$

$$+\cdots+\frac{S_2}{S_p}\frac{R_{2p}^{(1)}}{R_{22}^{(1)}}r_{1p}\,S_1S_p$$

$$= \frac{S_1S_2}{R_{22}^{(1)}}[r_{12}\,R_{22}^{(1)}+r_{13}\,R_{23}^{(1)}+r_{14}R_{24}^{(1)}+\cdots+r_{1p}\,R_{2p}^{(1)}]$$

$$= \frac{-S_1S_2}{R_{22}^{(1)}}R_{12}.$$

সুতরাং $r_{12\cdot34\cdots p} = \dfrac{-S_1S_2}{R_{22}^{(1)}}R_{12}\sqrt{\dfrac{R_{11}^{(2)}}{R^{(2)}}}\sqrt{\dfrac{R_{22}^{(1)}}{R^{(1)}}}\cdot\dfrac{1}{S_1S_2}$

অর্থাৎ $r_{12\cdot34p} = -\dfrac{R_{12}}{\sqrt{R_{11}R_{22}}}$, কারণ, $R_{22}^{(1)} = R_{11}^{(2)}$,

$R^{(2)} = R_{22}$ এবং $R^{(1)} = R_{11}$. $\cdots$ (11.13)

## 11.5 বহুল ও আংশিক সহগতি সম্পর্কে কয়েকটি তথ্য (Some facts concerning multiple and partial correlation) :

আমরা আগে যা দেখেছি তা থেকে পাওয়া যায় :—

1. $X_1 = X_{1\cdot23\ldots p}+x_{1\cdot23\ldots p}$

$= c+b_{12\cdot34\ldots p}X_2+\cdots+b_{1p\cdot23\ldots(p-1)}X_p+x_{1\cdot23\ldots p}$ ;

এখানে $b_{1j\cdot23\ldots(j-1)(j+1)\ldots p} = -\dfrac{S_1}{S_j}\dfrac{R_{1j}}{R_{11}}$, $(j=2,3,\ldots,p)$

2. $X_1 = X_{1\cdot34\ldots p}+x_{1\cdot34\ldots p}$

$= c+b_{13\cdot4\ldots p}X_3+b_{14\cdot35\ldots p}X_4+\cdots$

$+b_{1p\cdot34\ldots p-1}X_p+x_{1\cdot34\ldots p}$ ;

এখানে, $b_{1j\cdot34\ldots(j-1)(j+1)\ldots p} = -\dfrac{S_1}{S_j}\dfrac{R_{1j}^{(2)}}{R_{11}^{(2)}}$, $(j=3,4,\ldots,p)$ ;

3. $X_2 = X_{2\cdot34\ldots p} + x_{2\cdot34\ldots p}$

$$= d + b_{23\cdot4\ldots p}X_3 + b_{24\cdot35\ldots p}X_4$$

$$+ b_{2p\cdot34\ldots\overline{p-1}}X_p + x_{2\cdot34\ldots p}\,;$$

এখানে $b_{2j\cdot34\ldots(j-1)(j+1)\ldots p} = -\dfrac{S_2}{S_j}\dfrac{R_{2j}^{(1)}}{R_{22}^{(1)}}$, $(j = 3, 4, \ldots, P)$.

4. $r_{12\cdot34\ldots p} = \dfrac{\text{cov}\,(x_{1\cdot34\ldots p},\, x_{2\cdot34\ldots p})}{\sqrt{V(x_{1\cdot34\ldots p})}.\sqrt{V(x_{2\cdot34\ldots p})}} = \dfrac{-R_{12}}{\sqrt{R_{11}}\sqrt{R_{22}}}$.

5. $S^2{}_{1\cdot23\ldots p} = V(x_{1\cdot23\ldots p}) = S_1{}^2\,\dfrac{R}{R_{11}}$,

6. $S^2{}_{2\cdot13\ldots p} = V(x_{2\cdot134\ldots p}) = S_2{}^2\,\dfrac{R}{R_{22}}$.

7. $S_{1\cdot34\ldots p} = S_1\sqrt{\dfrac{R^{(2)}}{R_{11}{}^{(2)}}}$.

8. $S_{2\cdot34\ldots p} = S_2\sqrt{\dfrac{R^{(1)}}{R_{22}{}^{(1)}}}$.

9. $\dfrac{S_{1\cdot34\ldots p}}{S_{2\cdot34\ldots p}} = \dfrac{S_1}{S_2}\sqrt{\dfrac{R_{22}}{R_{11}}}$,

কারণ, $R^{(1)} = R_{11}$, $R^{(2)} = R_{22}$ এবং $R_{11}{}^{(2)} = R_{22}{}^{(1)}$

কাজেই আমরা লিখতে পারি

$$r_{12\cdot34\ldots p}\frac{S_{1\cdot23\ldots p}}{S_{2\cdot13\ldots p}} = -\frac{R_{12}}{\sqrt{R_{11}R_{22}}}\cdot\frac{S_1\sqrt{\dfrac{R}{R_{11}}}}{S_2\sqrt{\dfrac{R}{R_{22}}}}$$

$$= -\frac{R_{12}}{R_{11}}\frac{S_1}{S_2} = b_{12\cdot34\ldots p} \qquad \cdots \quad (11.14)$$

$$r_{12\cdot34\ldots p}\frac{S_{1\cdot34\ldots p}}{S_{2\cdot34\ldots p}} = -\frac{R_{12}}{\sqrt{R_{11}R_{22}}}\cdot\frac{S_1}{S_2}\cdot\sqrt{\frac{R_{22}}{R_{11}}}$$

$$= -\frac{R_{12}}{R_{11}}\frac{S_1}{S_2} = b_{12\cdot34\ldots p} \qquad \cdots \quad (11.15)$$

ফলে, $b_{12\cdot34\ldots p} = \dfrac{r_{12\cdot34\ldots p}S_{1\cdot34\ldots p}S_{2\cdot34\ldots p}}{V(x_{2\cdot34\ldots p})}$

$$= \frac{\text{cov}\,(x_{1\cdot34\ldots p},\, x_{2\cdot34\ldots p})}{V(x_{2\cdot34\ldots p})} \qquad \cdots \quad (11.16)$$

[ তুলনীয় $b_{12} = \dfrac{\text{cov}\,(X_1, X_2)}{V(X_2)}$ ]

স্পষ্টতঃই লেখা যাবে যে, $b_{21\cdot34\ldots p} = -\frac{R_{21}}{R_{22}}\cdot\frac{S_2}{S_1}$.

ফলে, $$b_{12\cdot34\ldots p}\times b_{21\cdot34\ldots p} = \left(-\frac{R_{21}}{R_{21}}\cdot\frac{S_2}{S_1}\right)\left(-\frac{R_{12}}{R_{11}}\cdot\frac{S_1}{S_2}\right)$$

$$= \left(-\frac{R_{12}}{\sqrt{R_{11}R_{22}}}\right)^2 = r^2_{12\cdot34\ldots p} \quad \cdots \quad (11.17)$$

[ তুলনীয় $b_{xy}\times b_{yx} = r_{xy}^2$ ]

আমরা আরও লিখতে পারি,

$$V(x_{1\cdot23\ldots p}) = \frac{1}{n}\sum x^2_{1\cdot23\ldots p\alpha}.$$

$$= \frac{1}{n}\sum_\alpha x_{1\cdot23\ldots p\alpha}\times$$

$$[X_{1\alpha} - b_{12\cdot34\ldots p}X_{2\alpha} - b_{13\cdot24\ldots p}X_{3\alpha}\ldots b_{1p\cdot23\ldots\overline{p-1}}X_{p\alpha}]$$

$$= \frac{1}{n}\sum_\alpha x_{1\cdot23\ldots p\alpha}X_{1\alpha}$$

[ নর্ম্যাল সমীকরণ ব্যবহার ক'রে ]

$$= \frac{1}{n}\sum_\alpha x_{1\cdot23\ldots p\alpha}\times$$

$$[X_{1\alpha} - b_{12\cdot34\ldots p}X_{2\alpha} - b_{13\cdot24\ldots p}X_{3\alpha}\ldots - b_{1p\cdot23\ldots\overline{p-1}}X_{p\alpha}]$$

[ নর্ম্যাল সমীকরণগুলি স্মরণে রেখে ]

$$= \frac{1}{n}\sum_\alpha x_{1\cdot23\ldots p\alpha}\times$$

$$[X_{1\alpha} - b_{12\cdot34\ldots p-1}X_{2\alpha} - b_{13\cdot24\ldots(p-1)}X_{3\alpha} - \cdots - b_{1(p-1)\cdot23\ldots(p-2)}X_{(p-1)\alpha}]$$

[ নর্ম্যাল সমীকরণগুলি স্মরণে রেখে ]

$$= \frac{1}{n}\sum x_{1\cdot23\ldots p\alpha}\cdot x_{1\cdot23\ldots(p-1)\alpha}$$

$$= \frac{1}{n}\sum x_{1\cdot23\ldots\overline{p-1}\alpha}[X_{1\alpha} - b_{12\cdot34\ldots p}X_{2\alpha} - b_{13\cdot24}$$

$$\ldots_p X_{3\alpha} - \ldots - b_{1(p-1)\cdot23\ldots(p-2)p}X_{(p-1)\alpha} - b_{1p\cdot23\ldots(p-1)}X_{p\alpha}]$$

$$= \frac{1}{n}\sum X_{1\alpha}\cdot x_{1\cdot 23\ldots(p-1)\alpha} - b_{1p\cdot 23\ldots(p-1)}$$

$$\frac{1}{n}\sum X_{p\alpha}\cdot x_{1\cdot 23\ldots(p-1)\alpha}$$

[ নর্ম্যাল সমীকরণগুলি স্মরণে রেখে ]

$$= \frac{1}{n}\sum x_{1\cdot 23\ldots\overline{p-1}\alpha}$$

$$[X_1 - b_{12\cdot 34\ldots(p-1)}X_{2\alpha} - b_{13\cdot 24\ldots(p-1)}X_{3\alpha} - \ldots - b_{1(p-1)\cdot 23\ldots(p-2)}X_{(p-1)\alpha}]$$

$$- b_{1p\cdot 23\ldots\overline{p-1}}\frac{1}{n}\sum x_{1\cdot 23\ldots(p-1)\alpha}$$

$$[X_{p\alpha} - b_{p2\cdot 34\ldots\overline{(p-1)}}X_{2\alpha} - \cdots b_{p\overline{p-1}\cdot 23\ldots(p-2)}X_{(p-1)\alpha}]$$

[ নর্ম্যাল সমীকরণ ব্যবহার ক'রে ]

$$= \frac{1}{n}\sum x^2_{1\cdot 23\ldots(p-1)} - b_{1p\cdot 23\ldots(p-1)}$$

$$\frac{1}{n}\sum x_{1\cdot 23\ldots(p-1)\alpha}x_{p\cdot 23\ldots(p-1)\alpha}$$

অর্থাৎ $S^2_{1\cdot 23\ldots p} = S^2_{1\cdot 23\ldots(p-1)} - b_{1p\cdot 23\ldots(p-1)}$

$\text{cov}\,(x_{1\cdot 23\ldots(p-1)}, x_{p\cdot 23\ldots(p-1)})$

$$= S^2_{1\cdot 23\ldots(p-1)} - b_{1p\cdot 23\ldots(p-1)}b_{p1\cdot 23\ldots(p-1)}S^2_{1\cdot 23\ldots(p-1)}$$

$$= S^2_{1\cdot 23\ldots(p-1)}(1 - b_{1p\cdot 23\ldots\overline{(p-1)}}b_{p1\cdot 23\ldots(p-1)})$$

$$= S^2_{1\cdot 23\ldots(p-1)}(1 - r^2_{1p\cdot 23\ldots\overline{(p-1)}}) \quad\cdots\quad (11.18)$$

সুতরাং $S^2_{1\cdot 23\ldots p} < S^2_{1\cdot 23\ldots\overline{(p-1)}}$ ··· (11.19)

অর্থাৎ $S_1^2(1 - r^2_{1\cdot 23\ldots p}) < S_1^2(1 - r^2_{1\cdot 23\ldots\overline{(p-1)}})$

অর্থাৎ $r^2_{1\cdot 23\ldots\overline{(p-1)}} < r^2_{1\cdot 23\ldots p}$ ··· ··· (11.20)

(11.18)-এ উল্লিখিত সূত্রটি বারবার ব্যবহার ক'রে পাওয়া যায়

$$S^2_{1,23\ldots p} = S^2_{1\cdot 23\ldots\overline{p-1}}(1 - r^2_{1p\cdot 23\ldots(p-1)})$$

$$= S^2_{1\cdot 23\ldots\overline{(p-2)}}(1 - r^2_{1\overline{(p-1)}\cdot 23\ldots\overline{(p-2)}})(1 - r^2_{1p\cdot 23\ldots\overline{(p-1)}})$$

$$= \cdots = S_1^2(1 - r_{12}^2)(1 - r^2_{13\cdot 2})(1 - r^2_{14\cdot 23}) \cdots (1 - r^2_{1p\cdot 23\ldots\overline{(p-2)}})$$

অর্থাৎ $1 - r^2_{1\cdot 23\ldots p} = (1 - r_{12}^2)(1 - r^2_{13\cdot 2})(1 - r^2_{14\cdot 23}) \cdots (1 - r^2_{1p\cdot 23\ldots(p-1)})$ ··· (11.21)

এই (11.21) সূত্রটি থেকে বহুল সহগাঙ্ক সহজেই নির্ণয় করা যায়।

[ যখন $p=4$,

তখন $r^2_{1\cdot 234}=1-(1-r^2_{12})(1-r^2_{13\cdot 2})(1-r^2_{14\cdot 23})$]

এর থেকে পাই

(1) $1-r^2_{1\cdot 23\ldots p} < 1-r^2_{12}$ অর্থাৎ $r^2_{1\cdot 23\ldots p} > r^2_{12}$

(2) $1-r^2_{1\cdot 23\ldots p} < 1-r^2_{13\cdot 2}$ অর্থাৎ $r^2_{1\cdot 23\ldots p} > r^2_{13\cdot 2}$

(3) $1-r^2_{1\cdot 23\ldots p} < 1-r^2_{14\cdot 23}$ অর্থাৎ $r^2_{1\cdot 23\ldots p} > r^2_{14\cdot 23}$

⋮

($p$) $1-r^2_{1\cdot 23\ldots p} < 1-r^2_{1p\cdot 23\ldots \overline{p-1}}$

অর্থাৎ $r^2_{1\cdot 23\ldots p} > r^2_{1p\cdot 23\ldots \overline{p-1}}$ অর্থাৎ বহুল সহগাঙ্কের মান কোন আংশিক বা পূর্ণ সহগাঙ্কের মানের চেয়ে কখনই কম হতে পারে না।

$$\text{cov}\,(x_{1\cdot 34\ldots p}, x_{2\cdot 34\ldots p})=\frac{1}{n}\sum x_{1\cdot 34\ldots p}x_{2\cdot 34\ldots p}$$

$$=\frac{1}{n}\sum x_{1\cdot 34\ldots p}\,[X_2-b_{23\cdot 4\ldots p}X_3-b_{24\cdot 35\ldots p}X_4-\cdots-b_{2p\cdot 34\ldots(\overline{p-1})}X_p]$$

$$=\frac{1}{n}\sum x_{1\cdot 34\cdots p}X_2=\frac{1}{n}\sum x_{1\cdot 34\ldots p}\,[X_2-b_{23\cdot 4\ldots(\overline{p-1})}X_3-\cdots-b_{2(\overline{p-1})\cdot 34\ldots(\overline{p-2})}X_{p-1}]$$

$$=\frac{1}{n}\sum x_{2\cdot 34\ldots\overline{p-1}}x_{1\cdot 34\ldots p}$$

$$=\frac{1}{n}\sum x_{2\cdot 34\ldots\overline{p-1}}[X_1-c-b_{13\cdot 4\ldots p}X_3-\cdots-b_{1p\cdot 34\ldots(\overline{p-1})}X_p]$$

$$=\frac{1}{n}\sum x_{2\cdot 34\ldots\overline{p-1}}[X_1-c-b_{1p\cdot 34\ldots(\overline{p-1})}X_p]$$

$$=\frac{1}{n}\sum X_1x_{2\cdot 34\ldots\overline{p-1}}-b_{1p\cdot 34\ldots\overline{p-1}}\frac{1}{n}\sum X_p x_{2\cdot 34\ldots(\overline{p-1})}$$

$$=\frac{1}{n}\sum x_{2\cdot 34\ldots(\overline{p-1})}[X_1-A-b_{13\cdot 4\ldots(\overline{p-1})}X_3-\cdots-b_{1\overline{p-1}\cdot 34\ldots(\overline{p-2})}X_{p-1}]$$

$$-b_{1p\cdot 34\ldots(\overline{p-1})}\frac{1}{n}\sum x_{2\cdot 34\ldots(\overline{p-1})}[X_p-B-b_{p3\cdot 4\ldots(\overline{p-1})}X_3-\cdots-b_{p(\overline{p-1})\cdot 34\ldots(\overline{p-2})}X_{p-1}]$$

$$= \frac{1}{n}\sum x_{1\cdot 34\ldots\overline{p-1}}x_{2\cdot 34\ldots\overline{p-1}} - b_{1p\cdot 34\ldots(\overline{p-1})}$$

$$\frac{1}{n}\sum x_{2\cdot 34\ldots(\overline{p-1})}\, x_{p\cdot 34\ldots\overline{p-1}}$$

সুতরাং $b_{12\cdot 34\ldots p}\, S^2_{2\cdot 34\ldots p} = b_{12\cdot 34\ldots\overline{p-1}} S^2_{2\cdot 34\ldots\overline{p-1}} - b_{1p\cdot 34\ldots\overline{p-1}}\, b_{p2\cdot 34\ldots\overline{p-1}}\, S^2_{2\cdot 34\ldots\overline{p-1}}$

অথবা $b_{12\cdot 34\ldots p} = (b_{12\cdot 34\ldots(\overline{p-1})} - b_{1p\cdot 34\ldots(\overline{p-1})}\, b_{p2\cdot 34\ldots(\overline{p-1})}) \dfrac{S^2_{2\cdot 34\ldots\overline{p-1}}}{S^2_{2\cdot 34\ldots p}}$

কিন্তু $S^2_{2\cdot 34\ldots p} = S^2_{2\cdot 34\ldots(\overline{p-1})}(1 - r^2_{2p\cdot 34\ldots(\overline{p-1})})$

$$= S^2_{2\cdot 34\ldots(\overline{p-1})}(1 - b_{2p\cdot 34\ldots(\overline{p-1})} b_{p2\cdot 34\ldots(\overline{p-1})}).$$

সুতরাং $b_{12\cdot 34\ldots p} = \dfrac{b_{12\cdot 34\ldots(\overline{p-1})} - b_{1p\cdot 34\ldots(\overline{p-1})} b_{p2\cdot 34\ldots(\overline{p-1})}}{1 - b_{2p\cdot 34\ldots(\overline{p-1})} b_{p2\cdot 34\ldots(\overline{p-1})}}$ ... (11.22)

[ যখন $p=4$, তখন $b_{12\cdot 34} = \dfrac{b_{12\cdot 3} - b_{14\cdot 3} b_{42\cdot 3}}{1 - b_{24\cdot 3} b_{42\cdot 3}}$ ]

কিন্তু $b_{12\cdot 34\ldots p} = r_{12\cdot 34\ldots p} \dfrac{S_{1\cdot 34\ldots p}}{S_{2\cdot 34\ldots p}}$.

$$b_{12\cdot 34\ldots(\overline{p-1})} = r_{12\cdot 34\ldots(\overline{p-1})} \frac{S_{1\cdot 34\ldots(\overline{p-1})}}{S_{2\cdot 34\ldots(\overline{p-1})}}$$

$$b_{1p\cdot 34\ldots(\overline{p-1})} = r_{1p\cdot 34\ldots(\overline{p-1})} \frac{S_{1\cdot 34\ldots(\overline{p-1})}}{S_{p\cdot 34\ldots(\overline{p-1})}}$$

$$b_{2p\cdot 34\ldots(\overline{p-1})} = r_{2p\cdot 34\ldots(\overline{p-1})} \frac{S_{2\cdot 34\ldots(\overline{p-1})}}{S_{p\cdot 34\ldots(\overline{p-1})}}$$

$$b_{p2\cdot 34\ldots(\overline{p-1})} = r_{2p\cdot 34\ldots\overline{p-1}} \frac{S_{p\cdot 34\ldots(\overline{p-1})}}{S_{2\cdot 34\ldots(\overline{p-1})}}$$

তাই $r_{12\cdot 34\ldots p} = \dfrac{S_{2\cdot 34\ldots p}}{S_{1\cdot 34\ldots p}}$

$$\times \frac{r_{12\cdot 34\ldots(\overline{p-1})} - r_{1p\cdot 34\ldots(\overline{p-1})} r_{2p\cdot 34\ldots(\overline{p-1})}}{(1 - r^2_{2p\cdot 34\ldots(\overline{p-1})})} \times \frac{S_{1\cdot 34\ldots(\overline{p-1})}}{S_{2\cdot 34\ldots(\overline{p-1})}}.$$

$$S^2_{2\cdot 34\ldots p} = S^2_{2\cdot 34\ldots(\overline{p-1})}(1 - r^2_{2p\cdot 34\ldots(\overline{p-1})})$$

এবং $S^2_{1\cdot 34\ldots p} = S^2_{1\cdot 34\ldots\overline{p-1}}(1 - r^2_{1p\cdot 34\ldots(\overline{p-1})})$

ফলে, $r_{12\cdot 34\ldots p} = \dfrac{r_{12\cdot 34\ldots(\overline{p-1})} - r_{1p\cdot 34\ldots\overline{p-1}} r_{2p\cdot 34\ldots(\overline{p-1})}}{\sqrt{(1 - r^2_{1p\cdot 34\ldots(\overline{p-1})})(1 - r^2_{2p\cdot 34\ldots(\overline{p-1})})}}$ ... (11.23)

যখন, $p=4$, তখন $r_{12\cdot34}=\dfrac{r_{12\cdot3}-r_{14\cdot3}r_{24\cdot3}}{\sqrt{(1-r^2_{24\cdot3})(1-r^2_{14\cdot3})}}$

$r_{12\cdot34\ldots p}$ কে বলা হয় $X_1$ ও $X_2$-এর $(p-2)$-ক্রমিক আংশিক সহগাঙ্ক, কারণ এতে $X_1$ ও $X_2$ ছাড়া অন্য $(p-2)$ সংখ্যক চল বিবেচনা করা হয়েছে; এর ক্রম বোঝানো হচ্ছে $r_{12\cdot34\ldots p}$ সংকেতসূচকে 2-এর পরবর্তী একটি বিন্দুর পরে ব্যবহৃত অঙ্কগুলির সংখ্যার সাহায্যে (3, 4,...$p$-মোট $(p-2)$-টি)। তেমনি $b_{12\cdot34\ldots p}$ হচ্ছে $X_2$-এর ওপর $X_1$-এর $(p-2)$ ক্রমিক আংশিক নির্ভরণাঙ্ক। ওপরের সমীকরণ দুটিতে [(11.22), (11.23)] দেখানো হচ্ছে কীভাবে কোন $(p-2)$ ক্রমিক সহগাঙ্ক ও নির্ভরণাঙ্ককে তাদের চেয়ে অধঃক্রমিক [ যথা $(p-3)$-ক্রমিক ] সহগাঙ্ক ও নির্ভরণাঙ্কের মাধ্যমে প্রকাশ করা যায়।

**উদাহরণ 11.1**

নিম্নলিখিত রাশিতথ্য থেকে গমের উৎপাদনের ওপর আবহাওয়া ও জলবায়ু সংক্রান্ত বিভিন্ন উপাদানের প্রভাব সম্পর্কে কিছু কিছু বিষয় জানা আছে। এই উপাদানগুলি কী তা নীচে বলা হয়েছে। উৎপাদনের ওপর এদের প্রভাব কী ভাবে মাপা যায় তা দেখ এবং $X_2$, $X_3$, $X_4$-এর মাধ্যমে $X_1$ সম্পর্কে পূর্বানুমানের উদ্দেশ্যে একটি পূর্বাভাষণসূত্র প্রতিষ্ঠা কর।

$X_1\equiv$ কোন উপযুক্ত এককে গমের গড় উৎপাদন ( মণে )

$X_2\equiv$ গত শীতঋতুতে বায়ুর গড় তাপ ( সেন্টিগ্রেডে )

$X_3\equiv$ প্রকৃত শস্যোৎপাদনকালে বায়ুর গড় তাপ ( সেন্টিগ্রেড )

$X_4\equiv$ শস্যোৎপাদন কালে মোট বৃষ্টিপাতের পরিমাণ ( সেন্টিগ্রেড )

হিসেবের সুবিধের জন্যে আমরা মূলবিন্দু ও মাপনা এককের পরিবর্তন ক'রে লিখব

$$u_1=\frac{X_1-2070}{10},\ u_2=(X_2-1\cdot8)\times10,$$

$$u_3=(X_3-12\cdot2)\times10 \text{ এবং } u_4=(X_4-278)$$

এখানে মোট পরিসংখ্যা হচ্ছে $n=30$.

তাহলে, $S_{ij}=\sum u_iu_j-\dfrac{(\Sigma u_i)(\Sigma u_j)}{n}$

এবং $\text{cor}\,(X_i, X_j)=\text{cor}\,(u_i, u_j)=\dfrac{S_{ij}}{\sqrt{S_{ii}}\sqrt{S_{jj}}}$.

**সারণী 11.1**

| বৎসর | $X_1$ | $X_2$ | $X_3$ | $X_4$ | বৎসর | $X_1$ | $X_2$ | $X_3$ | $X_4$ |
|---|---|---|---|---|---|---|---|---|---|
| 1913 | 1990 | 2·7 | 12·8 | 230 | 1928 | 2530 | 0·8 | 10·5 | 324 |
| 1914 | 1950 | 3·1 | 13·7 | 268 | 1929 | 2100 | 0·8 | 10·9 | 196 |
| 1915 | 1630 | 1·9 | 12·0 | 188 | 1930 | 2330 | 3·6 | 12·4 | 381 |
| 1916 | 1720 | 1·3 | 11·7 | 315 | 1931 | 1850 | 1·6 | 10·7 | 273 |
| 1917 | 1560 | 1·0 | 12·7 | 180 | 1932 | 2230 | 1·9 | 12·5 | 289 |
| 1918 | 1680 | 1·6 | 12·0 | 261 | 1933 | 2510 | 2·2 | 11·9 | 338 |
| 1919 | 1980 | 2·3 | 12·2 | 216 | 1934 | 2700 | 3·0 | 13·5 | 267 |
| 1920 | 2180 | 1·7 | 12·8 | 346 | 1935 | 2480 | 3·2 | 12·3 | 372 |
| 1921 | 2370 | 3·1 | 13·1 | 131 | 1936 | 1940 | 2·8 | 12·3 | 357 |
| 1922 | 1790 | 1·1 | 11·8 | 256 | 1937 | 2770 | 2·1 | 13·5 | 358 |
| 1923 | 2400 | 1·6 | 11·2 | 327 | 1938 | 2570 | 3·3 | 12·9 | 202 |
| 1924 | 1410 | 0·1 | 11·8 | 320 | 1939 | 2510 | 3·8 | 13·4 | 311 |
| 1925 | 2570 | 3·7 | 13·2 | 382 | 1940 | 1420 | −1·1 | 11·3 | 172 |
| 1926 | 2180 | 1·1 | 12·5 | 279 | 1941 | 810 | −0·4 | 11·3 | 194 |
| 1927 | 2150 | 2·5 | 12·2 | 351 | 1942 | 1990 | −2·4 | 11·2 | 261 |

$X_2$, $X_3$ ও $X_4$ অর্থাৎ $U_2$, $U_3$, ও $U_4$-এর মাধ্যমে $U_1$ অর্থাৎ $X_1$ সম্পর্কে লঘিষ্ঠবর্গনীতি অনুযায়ী নির্ণীত ঋজুরৈখিক নির্ভরণসূত্র সাহায্যে অনুমান করতে গিয়ে সূত্র পাওয়া যায়

$$U_1 = a + b_2u_2 + b_3u_3 + b_4u_4.$$

নির্ভরাঙ্ক নির্ণয়ে ব্যবহৃত নর্ম্যাল সমীকরণগুলি হ'ল

$$b_2S_{22} + b_3S_{32} + b_4S_{42} = S_{12}$$

$$b_2S_{23} + b_3S_{33} + b_4S_{43} = S_{13}$$

$$b_2S_{24} + b_3S_{34} + b_4S_{44} = S_{14}$$

এবং $a = \bar{u}_1 - b_2\bar{u}_2 - b_3\bar{u}_3 - b_4\bar{u}_4$

সূত্রটি শেষ পর্যন্ত দাঁড়ায়

$$\frac{X_1 - 2070}{10} = a + b_2(X_2 - 1·8) \times 10 + b_3(X_3 - 12·2) \times 10$$

$$+ b_4(X_4 - 278)$$

পূর্ণ সমাধান নির্ণয়ার্থে নীচের সারণিটি গঠন করতে হচ্ছে।

### সারণী 11.2

### বহুল ও আংশিক সহগাঙ্ক ও নির্ভরণাঙ্ক নির্ণয়

| বৎসর | $u_1$ | $u_2$ | $u_3$ | $u_4$ | $u_1u_2$ | $u_1u_3$ | $u_2u_3$ | $u_1u_4$ | $u_2u_4$ | $u_3u_4$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1913 | −8 | 9 | 6 | −48 | −72 | −48 | 54 | 384 | −432 | −288 |
| 1914 | −12 | 13 | 15 | −10 | −156 | −180 | 195 | 120 | −130 | −150 |
| 1915 | −44 | 1 | −2 | −90 | −44 | 88 | −2 | 3960 | −90 | 180 |
| 1916 | −35 | −5 | −5 | 37 | 175 | 175 | 25 | −1295 | −185 | −185 |
| 1917 | −51 | −8 | 5 | −98 | 408 | −255 | −40 | 4998 | 784 | −490 |
| 1918 | −39 | −2 | −2 | −17 | 78 | 78 | 4 | 663 | 34 | 34 |
| 1919 | −9 | 5 | 0 | −62 | −45 | 0 | 0 | 558 | −310 | 0 |
| 1920 | 11 | −1 | 6 | 68 | −11 | 66 | −6 | 748 | −68 | 408 |
| 1921 | 30 | 13 | 9 | −147 | 390 | 270 | 117 | −4410 | −1911 | 1323 |
| 1922 | −28 | −7 | −4 | −22 | 96 | 112 | 28 | 616 | 154 | 88 |
| 1923 | 33 | −2 | −10 | 49 | −65 | −330 | 20 | 1617 | −98 | −490 |
| 1924 | −66 | −17 | −4 | 42 | 1122 | 264 | 68 | −2772 | −7714 | −168 |
| 1925 | 50 | 19 | 10 | 104 | 950 | 500 | 190 | 5200 | 1976 | 1040 |
| 1926 | 11 | −7 | 3 | 1 | 77 | 33 | −21 | 11 | −7 | 3 |
| 1927 | 8 | 7 | 0 | 73 | 56 | 0 | 0 | 584 | 511 | 0 |
| 1928 | 46 | −10 | −17 | 46 | −460 | −782 | 170 | 2116 | −460 | −782 |
| 1929 | 3 | −10 | −13 | −82 | −30 | −39 | 130 | −246 | 820 | 1066 |
| 1930 | 26 | 18 | 2 | 103 | 468 | 52 | 36 | 2678 | 1854 | 206 |
| 1931 | −22 | −2 | −15 | −5 | 46 | 330 | 30 | 110 | 10 | 75 |
| 1932 | 16 | 1 | 3 | 11 | 16 | 48 | 3 | 176 | 11 | 33 |
| 1933 | 44 | 4 | −3 | 60 | −176 | −132 | −12 | 2640 | 240 | −180 |
| 1934 | 53 | 12 | 13 | −11 | 636 | 869 | 156 | −583 | −132 | −143 |
| 1935 | 41 | 14 | 1 | 94 | 574 | 41 | 14 | 3854 | 1316 | 94 |
| 1939 | −13 | 10 | 1 | 79 | −13 | −13 | 10 | −1027 | 790 | 79 |
| 1937 | 70 | 8 | 13 | 80 | 210 | 910 | 39 | 5600 | 240 | 1040 |
| 1938 | 50 | 15 | 7 | −76 | 750 | 350 | 105 | −3800 | −1140 | −532 |
| 1939 | 44 | 20 | 12 | 33 | 880 | 528 | 240 | 1452 | 660 | 396 |
| 1940 | −65 | −29 | −9 | −106 | 1885 | 585 | 261 | 6890 | 3079 | 954 |
| 1941 | −126 | −22 | −9 | −84 | 2172 | 1134 | 198 | 10584 | 1848 | 756 |
| 1942 | −8 | −42 | −10 | −17 | 336 | 80 | 420 | 136 | 714 | 170 |
| সমষ্টি | 10 | 0 | 3 | 5 | 11031 | 4554 | 2432 | 1562 | 9359 | 1891 |

এ ছাড়া আরও পাওয়া যায়

$$\Sigma u_1^2=57144,\ \Sigma u_2^2=6048,\ \Sigma u_3^2=2177,\ \Sigma u_4^2=142877.$$

$$\bar{u}_1=\cdot33333,\ \bar{u}_2=0,\ \bar{u}_3=\cdot100000,\ \bar{u}_4=\cdot166667$$

$$S_{11}=57140\cdot6667,\ S_{12}=11031,\ S_{13}=4553,$$

$$S_{14}=41560\cdot3333,\ S_{22}=6048,\ S_{23}=2432,\ S_{24}=9359,$$

$$S_{33}=2176\cdot7000,\ S_{34}=1890\cdot5,\ S_{44}=142876\cdot1667.$$

তাহলে পাওয়া যায়

$$b_2=1\cdot3567,\quad b_3=0\cdot4051,\ b_4=0\cdot1967,\ a=0\cdot2600$$

এবং পূর্বাভাষ সূত্রটি হ'ল

$$X_1=787\cdot4601+135\cdot6730\ X_2+40\cdot5113\ X_3+1\cdot9665\ X_4.$$

এছাড়া আরও পাওয়া যায়

$$r_{12}=0\cdot5934,\quad r_{13}=0\cdot4082,\quad r_{18}=0\cdot4600,$$

$$r_{23}=0\cdot6703,\quad r_{24}=0\cdot3184,\quad r_{34}=0\cdot1072,$$

$$r_{12\cdot3}=0\cdot4724,\ r_{13\cdot2}=0\cdot0176,\ r_{14\cdot3}=0\cdot4586,$$

$$r_{14\cdot2}=0\cdot372,\quad r_{24\cdot3}=0\cdot3341,\ r_{23\cdot4}=-0\cdot1510,$$

$$r_{12\cdot34}=0\cdot3811,\ r_{13\cdot24}=0\cdot0771,\ r_{14\cdot23}=0\cdot3620.$$

$$R^2_{1\cdot234}=0.4372,\ R_{1\cdot234}=0.6612.$$

এখন লক্ষণীয় যে, max $r_{1j}=r_{12}$-এর মান মোটামুটি বেশী। কাজেই $X_1$-এর ওপর $X_2$-এর প্রভাব প্রণিধানযোগ্য। তাছাড়া, max $r_{1j\cdot2}=r_{14\cdot2}$-এর পরিমাণও কম নয়। কাজেই $X_2$ ছাড়া $X_4$ও $X_1$-এর ওপর প্রভাব বিস্তার করে। সব শেষে $r_{13\cdot24}$-এর মান অবশ্য সামান্য। কাজেই $X_1$-এর ওপর $X_3$-এর প্রভাব তেমন কিছু নয়।

**টীকা।** বহুল সহগাঙ্ক নির্ণয়ে নিম্নলিখিত বিষয়টি অনুধাবনযোগ্য :

ধরা যাক $X_1, X_2, X_3,\ldots$ইত্যাদি হচ্ছে পরস্পর নিরপেক্ষ চল এবং $Y$ হচ্ছে তাদের ওপর নির্ভরশীল চল। কাজেই $X_1, X_2, X_3,\ldots$ এর ওপর $Y$-এর নির্ভরণ নির্ণয় করতে গিয়ে নির্ভরণ রেখা $\hat{Y}=a+b_1X_1+b_2X_2+b_3X_3+\cdots$ ব্যবহার করতে হয়। তাহলে, $a, b_1, b_2, b_3,\cdots$ ইত্যাদি নির্ণয়ে ব্যবহৃত নর্ম্যাল সমীকরণগুলি দাঁড়ায়

$$\sum Y = na + b_1 \sum X_1 + b_2 \sum X_2 + \cdots$$
$$\sum YX_1 = a \sum X_1 + b_1 \sum X_1^2 + b^2 \sum X_1 X_2 + \cdots$$
$$\sum YX_2 = a \sum X_2 + b_1 \sum X_1 X_2 + b_2 \sum X_2^2 + \cdots$$
$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

এখন যদি লেখা যায়, $S_{YY} = \sum (Y - \overline{Y})^2$,

$SY_i = \sum (Y - \overline{Y})(X_i - \overline{X}_i)$ ও $S_{ji} = S_{ij} = \sum (X_i - \overline{X}_i)(X_j - \overline{X}_j)$,

তাহলে নর্ম্যাল সমীকরণগুলি দাঁড়ায় ( প্রথমটি বাদ দিয়ে )

$$SY_1 = b_1 S_{11} + b_2 S_{21} + b_3 S_{31} + \cdots$$
$$SY_2 = b_1 S_{12} + b_2 S_{22} + b_3 S_{32} + \cdots$$
$$SY_3 = b_1 S_{13} + b_2 S_{23} + b_3 S_{33} + \cdots$$
$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

এগুলিকে সমাধান ক'রে ধর $b_i$-এর অনুমিত মান বের করা হয়েছে

$$\widehat{b}_i \ (i = 1, 2, \ldots).$$

এখন, $$r_{1 \cdot 234 \cdots} = \frac{\text{cov}(Y, \widehat{Y})}{\sqrt{V(Y)}\sqrt{V(\widehat{Y})}} \sqrt{\frac{\text{cov}(Y, \widehat{Y})}{V(Y)}}$$

$$= \sqrt{\frac{\sum (Y - \overline{Y})(\widehat{Y} - \overline{\widehat{Y}})}{\sum (Y - \overline{Y})^2}} = \sqrt{\frac{\sum (Y - \overline{Y})(\widehat{Y} - \overline{\widehat{Y}})}{S_{YY}}}$$

কিন্তু $$\sum (Y - \overline{Y})(\widehat{Y} - \overline{\widehat{Y}}) = \sum (Y - \overline{Y})$$
$$[\, b_1 (X_1 - \overline{X}_1) + b_2 (X_2 - \overline{X}_2) + \cdots]$$
$$= b_1 S_{Y1} + b_2 S_{Y2} + b_{Y3} S_{r3} + \cdots$$

এবং নর্ম্যাল সমীকরণ ব্যবহার ক'রে এর অনুমিত মান হচ্ছে

$$\widehat{b}_1 S_{Y1} + \widehat{b}_2 S_{Y2} + \widehat{b}_3 S_{Y3} + \cdots$$

সুতরাং $r_{1 \cdot 234} \ldots$.এর অনুমিত এবং ব্যবহার্য মান হচ্ছে

$$\sqrt{\frac{\widehat{b}_1 S_{Y1} + \widehat{b}_2 S_{2Y} + \cdots}{S_{YY}}} \qquad \cdots (11.22)$$

কোন নমুনার ভিত্তিতে $r_{1\cdot 23\ldots}$-এর মান নির্ণয়ে এই সূত্রের ব্যবহার সর্বোত্তম, অবশ্য যদি সেই সঙ্গে বিভিন্ন সহগাঙ্ক $r_{ij}\ (i, j=1, 2,\cdots)$ ইত্যাদিকে পৃথকভাবে নির্ণয় করার প্রয়োজন না হয়।

## অনুশীলনী

11.1 যদি $a_1 x_1 + a_2 x_2 + a_3 x_3 = c$ হয়,
তাহলে দেখাও যে, $\rho^2_{12\cdot 3} = \rho^2_{13\cdot 2} = \rho^2_{23\cdot 1} = 1.$

11.2. দেখাও যে

(i) $r_{1\cdot 23} = \sqrt{\dfrac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1-r_{23}^2}}$

(ii) $r_{12\cdot 3} = \dfrac{r_{12} - r_{13}r_{23}}{\sqrt{1-r_{13}^2}\,.\,\sqrt{1-r_{23}^2}}.$

(iii) $0 \leqslant r_{1\cdot 23\ldots p} \leqslant 1$

(iv) $-1 \leqslant r_{12\cdot 34\ldots p} \leqslant +1$

11.3. ধর $p$-সংখ্যক চল $X_1,\ldots X_p$-এর জন্যে দেওয়া আছে

$$\text{cor}\ (X_i, X_j) = r_{ij} = r (i, j = 1, 2,\cdots p,\ i \neq j).$$

তাহলে $r_{1\cdot 23\ldots p}$ ও $r_{12\cdot 34\ldots p}$-এর মান কত হবে?

$$\left[\text{উত্তর :}\quad \left[\frac{(p-1)r^2}{1+(p-2)r}\right]^{\frac{1}{2}},\ \frac{r}{1+(p-2)r}\right]$$

11.4. যদি $p$-সংখ্যক চল $X_1,\ldots, X_p$-এর জন্যে

$$\text{cor}\ (X_1, X_i) = r_{1i} = r\ (i = 2, 3,\cdots p)$$

এবং $\text{cor}\ (X_i, X_j) = r_{ij} = r'\ (i, j = 2, 3,\ldots p,\ i \neq j)$ হয়,

তাহলে $r_{1\cdot 23\ldots p}$ ও $r_{12\cdot 34\ldots p}$-এর মান কত হবে?

$$\left[\text{উত্তর :}\quad \left[\frac{r^2(p-1)}{1+(p-2)r'}\right]^{\frac{1}{2}},\ \frac{r}{1+(p-2)r'}\right]$$

11.5. তিনটি চল $X_1$ (মিলিমিটারে দৈর্ঘ্য), $X_2$ (ঘনসেন্টিমিটারে আয়তন) এবং $X_3$ (গ্রাম এককে ওজন)-এর সম্পর্কে 300টি ডিমের জন্যে পরিমাপ নেওয়া হয়েছে এবং তাদের সম্পর্কে গড়, প্রমাণবিচ্যুতি এবং সহগাঙ্ক বিষয়ে তথ্য নীচে দেওয়া রয়েছে :

$\bar{X}_1 = 55{\cdot}95$ মি. মি. | $S_1 = 2{\cdot}26$ মি. মি.
$\bar{X}_2 = 51{\cdot}48$ ঘনসেন্টিমিটার | $S_2 = 4{\cdot}39$ ঘন সে. মি.
$\bar{X}_3 = 56{\cdot}03$ গ্রাম | $S_3 = 4{\cdot}41$ গ্রাম

$$r_{12} = 0{\cdot}578,\ r_{13} = 0{\cdot}581,\ r_{23} = 0{\cdot}974$$

(*a*) দৈর্ঘ্য ও আয়তনের ওপর ওজনের বহুল নির্ভরণ রেখার সমীকরণ নির্ণয় কর এবং তার থেকে প্রদত্ত 58 মি. মি. দৈর্ঘ্য ও 57·5 ঘন সে. মি. আয়তনবিশিষ্ট একটি ডিমের ওজনের একটি সঙ্গত প্রাক্কলক নিরূপণ কর।

(*b*) ডিমের ওজন এবং আয়তনের উভয়ের ওপর দৈর্ঘ্যের প্রভাব বিদূরিত ক'রে তাদের সহগাঙ্ক নির্ণয় কর।

11.6. যদি $r_{12} = r_{13} = r_{23} = r$ হয়, তবে দেখাও যে, $r \geqslant -\frac{1}{2}$

[ উদাহরণ 11.3-এর ফল ব্যবহার কর ]

11.7 $x_1,\ x_2,\ x_3$-এর গড় যদি শূন্য হয় এবং $x_2$ ও $x_3$-এর ওপর $x_1$-এর নির্ভরণ সরলরেখা যদি

$\hat{x}_1 = b_{12\cdot3}\,x_2 + b_{13\cdot2}\,x_3$ হয় তবে $\hat{b}_{12\cdot3}$-কে লঘিষ্ঠ বর্গনীতি অনুযায়ী নির্ণীত $b_{12\cdot3}$-এর প্রাক্কলক লিখে একে $b_{12}$ ইত্যাদির মাধ্যমে প্রকাশ কর। আবার, $u = x_1 - b_{13}\,x_3$ ও $v = x_2 - b_{23}x_3$ লিখে $v$-এর ওপর $u$-এর নির্ভরণ রেখাকে $\hat{u} = b'v$ লিখে দেখাও যে, $b' = b_{12\cdot3}$.

11.8 তিনটি চল $X_1,\ X_2,\ X_3$ বিবেচনা ক'রে দেখাও যে এক্ষেত্রে সহগাঙ্ক ডিটারমিন্যান্ট $|R|$-এর মান অ-ঋণাত্মক হবে। আরও দেখাও যে,

$$\rho_{23}\ \varepsilon\ [\rho_{12}\,\rho_{13} \pm (1 - \rho_{12}^2 - \rho_{13}^2 + \rho_{12}^2\,\rho_{13})^{\frac{1}{2}}]$$

## নির্দেশিকা

1. Goon, A. M., Gupta, M. K. and Dasgupta, B. *Fundamentals of Statistics*, Vol. 1. The World Press, Pvt. Ltd., 1970.

2, Yule, G. U. and Kendall, M. G. *An introduction to the theory* of *Statistics*. Charles Griffin, 1950.

# সারণী 1

## মৌল নর্ম্যাল বিভাজনের কোটি এবং ক্ষেত্রফল*

| $\tau$ | $\phi(\tau)$ | $\Phi(\tau)$ | $\tau$ | $\phi(\tau)$ | $\Phi(\tau)$ | $\tau$ | $\phi(\tau)$ | $\Phi(\tau)$ |
|---|---|---|---|---|---|---|---|---|
| .00 | .3989423 | .5000000 | | | | | | |
| .01 | .3989223 | .5039894 | .51 | .3502919 | .6949743 | 1.01 | .2395511 | .8437524 |
| .02 | .3988625 | .5079783 | .52 | .3484925 | .6984682 | 1.02 | .2371320 | .8461358 |
| .03 | .3987628 | .5119665 | .53 | .3466677 | .7019440 | 1.03 | .2347138 | .8484950 |
| .04 | .3986233 | .5159534 | .54 | .3448180 | .7054015 | 1.04 | .2322970 | .8508300 |
| .05 | .3984439 | .5199388 | .55 | .3429439 | .7088403 | 1.05 | .2298821 | .8531409 |
| .06 | .3982248 | .5239222 | .56 | .3410458 | .7122603 | 1.06 | .2274696 | .8554277 |
| .07 | .3979661 | .5279032 | .57 | .3391243 | .7156612 | 1.07 | .2250599 | .8576903 |
| .08 | .3976677 | .5318814 | .58 | .3371799 | .7190427 | 1.08 | .2226535 | .8599289 |
| .09 | .3973298 | .5358564 | .59 | .3352132 | .7224047 | 1.09 | .2202508 | .8621434 |
| .10 | .3969525 | .5398278 | .60 | .3332246 | .7257469 | 1.10 | .2178522 | .8643339 |
| .11 | .3965360 | .5437953 | .61 | .3312147 | .7290691 | 1.11 | .2154582 | .8665005 |
| .12 | .3960802 | .5477584 | .62 | .3291840 | .7323711 | 1.12 | .2130691 | .8686431 |
| .13 | .3955854 | .5517168 | .63 | .3271330 | .7356527 | 1.13 | .2106856 | .8707619 |
| .14 | .3950517 | .5556700 | .64 | .3250623 | .7389137 | 1.14 | .2083078 | .8728568 |
| .15 | .3944793 | .5596177 | .65 | .3229724 | .7421539 | 1.15 | .2059363 | .8749281 |
| .16 | .3938684 | .5635595 | .66 | .3208638 | .7453731 | 1.16 | .2035714 | .8769756 |
| .17 | .3932190 | .5674949 | .67 | .3187371 | .7485711 | 1.17 | .2012135 | .8789995 |
| .18 | .3925315 | .5714237 | .68 | .3165929 | .7517478 | 1.18 | .1988631 | .8809999 |
| .19 | .3918060 | .5753454 | .69 | .3144317 | .7549029 | 1.19 | .1965205 | .8829768 |
| .20 | .3910427 | .5792597 | .70 | .3122539 | .7580363 | 1.20 | .1941861 | .8849303 |
| .21 | .3902419 | .5831662 | .71 | .3100603 | .7611479 | 1.21 | .1918602 | .8868606 |
| .22 | .3894038 | .5870644 | .72 | .3078513 | .7642375 | 1.22 | .1895432 | .8887676 |
| .23 | .3885286 | .5909541 | .73 | .3056274 | .7673049 | 1.23 | .1872354 | .8906514 |
| .24 | .3876166 | .5948349 | .74 | .3033893 | .7703500 | 1.24 | .1849373 | .8925123 |
| .25 | .3866681 | .5987063 | .75 | .3011374 | .7733726 | 1.25 | .1826491 | .8943502 |
| .26 | .3856834 | .6025681 | .76 | .2988724 | .7763727 | 1.26 | .1803712 | .8961653 |
| .27 | .3846627 | .6064199 | .77 | .2965948 | .7793501 | 1.27 | .1781038 | .8979577 |
| .28 | .3836063 | .6102612 | .78 | .2943050 | .7823046 | 1.28 | .1758474 | .8997274 |
| .29 | .3825146 | .6140919 | .79 | .2920038 | .7852361 | 1.29 | .1736022 | .9014747 |
| .30 | .3813878 | .6179114 | .80 | .2896916 | .7881446 | 1.30 | .1713686 | .9031995 |
| .31 | .3802264 | .6217195 | .81 | .2873689 | .7910299 | 1.31 | .1691468 | .9049021 |
| .32 | .3790305 | .6255158 | .82 | .2850364 | .7938919 | 1.32 | .1669370 | .9065825 |
| .33 | .3778007 | .6293000 | .83 | .2826945 | .7967306 | 1.33 | .1647397 | .9082409 |
| .34 | .3765372 | .6330717 | .84 | .2803438 | .7995458 | 1.34 | .1625551 | .9098773 |
| .35 | .3752403 | .6368307 | .85 | .2779849 | .8023375 | 1.35 | .1603833 | .9114920 |
| .36 | .3739106 | .6405764 | .86 | .2756182 | .8051055 | 1.36 | .1582248 | .9130850 |
| .37 | .3725483 | .6443088 | .87 | .2732444 | .8078498 | 1.37 | .1560797 | .9146565 |
| .38 | .3711539 | .6480273 | .88 | .2708640 | .8105703 | 1.38 | .1539483 | .9162067 |
| .39 | .3697277 | .6517317 | .89 | .2684774 | .8132671 | 1.39 | .1518308 | .9177356 |
| .40 | .3682701 | .6554217 | .90 | .2660852 | .8159399 | 1.40 | .1497275 | .9192433 |
| .41 | .3667817 | .6590970 | .91 | .2636880 | .8185887 | 1.41 | .1476385 | .9207302 |
| .42 | .3652627 | .6627573 | .92 | .2612863 | .8212136 | 1.42 | .1455641 | .9221962 |
| .43 | .3637136 | .6664022 | .93 | .2588805 | .8238145 | 1.43 | .1435046 | .9236415 |
| .44 | .3621349 | .6700314 | .94 | .2564713 | .8263912 | 1.44 | .1414600 | .9250663 |
| .45 | .3605270 | .6736448 | .95 | .2540591 | .8289439 | 1.45 | .1394306 | .9264707 |
| .46 | .3588903 | .6772419 | .96 | .2516443 | .8314724 | 1.46 | .1374165 | .9278505 |
| .47 | .3572253 | .6808225 | .97 | .2492277 | .8339768 | 1.47 | .1354181 | .9292191 |
| .48 | .3555325 | .6843863 | .98 | .2468095 | .8364569 | 1.48 | .1334353 | .9305634 |
| .49 | .3538124 | .6879331 | .99 | .2443904 | .8389129 | 1.49 | .1314684 | .9318879 |
| .50 | .3520653 | .6914625 | 1.00 | .2419707 | .8413447 | 1.50 | .1295176 | .9331928 |

## সারণী 1 ( পূর্বানুবৃত্ত )

| ব | $\phi$(ব) | $\Phi$(ব) | ব | $\phi$(ব) | $\Phi$(ব) | ব | $\phi$(ব) | $\Phi$(ব) |
|---|---|---|---|---|---|---|---|---|
| 1.51 | .1275830 | .9344783 | 2.01 | .0529192 | .9777844 | 2.51 | .0170947 | .9939634 |
| 1.52 | .1256646 | .9357445 | 2.02 | .0518636 | .9783083 | 2.52 | .0166701 | .9941323 |
| 1.53 | .1237628 | .9369916 | 2.03 | .0508239 | .9788217 | 2.53 | .0162545 | .9942969 |
| 1.54 | .1218775 | .9382198 | 2.04 | .0498001 | .9793248 | 2.54 | .0158476 | .9944574 |
| 1.55 | .1200090 | .9394292 | 2.05 | .0487920 | .9798178 | 2.55 | .0154493 | .9946139 |
| 1.56 | .1181573 | .9406201 | 2.06 | .0477996 | .9803007 | 2.56 | .0150596 | .9947664 |
| 1.57 | .1163225 | .9417924 | 2.07 | .0468226 | .9807738 | 2.57 | .0146782 | .9949151 |
| 1.58 | .1145048 | .9429466 | 2.08 | .0458611 | .9812372 | 2.58 | .0143051 | .9950600 |
| 1.59 | .1127042 | .9440826 | 2.09 | .0449148 | .9816911 | 2.59 | .0139401 | .9952012 |
| 1.60 | .1109208 | .9452007 | 2.10 | .0439836 | .9821356 | 2.60 | .0135830 | .9953388 |
| 1.61 | 1091548 | .9463011 | 2.11 | .0430674 | .9825708 | 2.61 | .0132337 | .9954729 |
| 1.62 | .1074061 | .9473839 | 2.12 | .0421661 | .9829970 | 2.62 | .0128921 | .9956035 |
| 1.63 | 1056748 | .9484493 | 2.13 | .0412795 | .9834142 | 2.63 | .0125581 | .9957308 |
| 1.64 | .1039611 | .9494974 | 2.14 | .0404076 | .9838226 | 2.64 | .0122315 | .9958547 |
| 1.65 | .1022649 | .9505285 | 2.15 | .0395500 | .9842224 | 2.65 | .0119122 | .9959754 |
| 1.66 | .1005864 | .9515428 | 2.16 | .0387069 | .9846137 | 2.66 | .0116001 | .9960930 |
| 1.67 | .0989255 | .9525403 | 2.17 | .0378779 | .9849966 | 2.67 | .0112951 | .9962074 |
| 1.68 | .0972823 | .9535213 | 2.18 | .0370629 | .9853713 | 2.68 | .0109969 | .9963189 |
| 1.69 | .0956568 | .9544860 | 2.19 | .0362619 | .9857379 | 2.69 | .0107056 | .9964274 |
| 1.70 | .0940491 | .9554345 | 2.20 | .0354746 | .9860966 | 2.70 | .0104209 | .9965330 |
| 1.71 | .0924591 | .9563671 | 2.21 | .0347009 | .9864474 | 2.71 | .0101428 | .9966358 |
| 1.72 | .0908870 | .9572838 | 2.22 | .0339408 | .9867906 | 2.72 | .0098712 | .9967359 |
| 1.73 | .0893326 | .9581849 | 2.23 | .0331939 | .9871263 | 2.73 | .0096058 | .9968333 |
| 1.74 | .0877961 | .9590705 | 2.24 | .0324603 | .9874545 | 2.74 | .0093466 | .9969280 |
| 1.75 | .0862773 | .9599408 | 2.25 | .0317397 | .9877755 | 2.75 | .0090936 | .9970202 |
| 1.76 | .0847764 | .9607961 | 2.26 | .0310319 | .9880894 | 2.76 | .0088465 | .9971099 |
| 1.77 | .0832932 | .9616364 | 2.27 | .0303370 | .9883962 | 2.77 | .0086052 | .9971972 |
| 1.78 | .0818278 | .9624620 | 2.28 | .0296546 | .9886962 | 2.78 | .0083697 | .9972821 |
| 1.79 | .0803801 | .9632730 | 2.29 | .0289847 | .9889893 | 2.79 | .0081398 | .9973646 |
| 1.80 | .0789502 | .9640697 | 2.30 | .0283270 | .9892759 | 2.80 | .0079155 | .9974449 |
| 1.81 | .0775379 | .9648521 | 2.31 | .0276816 | .9895559 | 2.81 | .0076965 | .9975229 |
| 1.82 | .0761433 | .9656205 | 2.32 | .0270481 | .9898296 | 2.82 | 0074829 | .9975988 |
| 1.83 | .0747663 | .9663750 | 2.33 | .0264265 | .9900969 | 2.83 | .0072744 | .9976726 |
| 1.84 | .0734068 | .9671159 | 2.34 | .0258166 | .9903581 | 2.84 | .0070711 | .9977443 |
| 1.85 | .0720649 | .9678432 | 2.35 | 0252182 | .9906133 | 2.85 | .0068728 | .9978140 |
| 1.86 | .0707404 | .9685572 | 2.36 | .0246313 | .9908625 | 2.86 | .0066793 | .9978818 |
| 1.87 | .0694333 | .9692581 | 2.37 | .0240556 | .9911060 | 2.87 | .0064907 | .9979476 |
| 1.88 | .0681436 | .9699460 | 2.38 | .0234910 | .9913437 | 2.88 | .0063067 | .9980116 |
| 1.89 | .0668711 | .9706210 | 2.39 | .0229374 | .9915758 | 2.89 | .0061274 | .9980738 |
| 1.90 | .0656158 | .9712834 | 2.40 | .0223945 | .9918025 | 2.90 | .0059525 | .9981342 |
| 1.91 | .0643777 | .9719334 | 2.41 | .0218624 | .9920237 | 2.91 | .0057821 | .9981929 |
| 1.92 | .0631566 | .9725711 | 2.42 | .0213407 | .9922397 | 2.92 | .0056160 | .9982498 |
| 1.93 | .0619524 | .9731966 | 2.43 | .0208294 | .9924506 | 2.93 | .0054541 | .9983052 |
| 1.94 | .0607652 | .9738102 | 2.44 | .0203284 | .9926564 | 2.94 | .0052963 | .9983589 |
| 1.95 | .0595947 | .9744119 | 2.45 | .0198374 | .9928572 | 2.95 | .0051426 | .9984111 |
| 1.96 | .0584409 | .9750021 | 2.46 | .0193563 | .9930531 | 2.96 | .0049929 | .9984618 |
| 1.97 | .0573038 | .9755808 | 2.47 | .0188850 | .9932443 | 2.97 | .0048470 | .9985110 |
| 1.98 | .0561831 | .9761482 | 2.48 | .0184233 | .9934309 | 2.98 | .0047050 | .9985588 |
| 1.99 | .0550789 | .9767045 | 2.49 | .0179711 | .9936128 | 2.99 | .0045666 | .9986051 |
| 2.00 | .0539910 | .9772499 | 2.50 | .0175283 | .9937903 | 3.00 | .0044318 | .9986501 |

## সারণী 1 ( পূর্বানুবৃত্ত )

| $\tau$ | $\phi(\tau)$ | $\Phi(\tau)$ | $\tau$ | $\phi(\tau)$ | $\Phi(\tau)$ | $\tau$ | $\phi(\tau)$ | $\Phi(\tau)$ |
|---|---|---|---|---|---|---|---|---|
| 3.01 | .0043007 | .9986938 | 3.21 | .0023089 | .9993363 | 3.41 | .0011910 | .9996752 |
| 3.02 | .0041729 | .9987361 | 3.22 | .0022358 | .9993590 | 3.42 | .0011510 | .9996869 |
| 3.03 | .0040486 | .9987772 | 3.23 | .0021649 | .9993810 | 3.43 | .0011122 | .9996982 |
| 3.04 | .0039276 | .9988171 | 3.24 | .0020960 | .9994024 | 3.44 | .0010747 | .9997091 |
| 3.05 | .0038098 | .9988558 | 3.25 | .0020290 | .9994230 | 3.45 | .0010383 | .9997197 |
| 3.06 | .0036951 | .9988933 | 3.26 | .0019641 | .9994429 | 3.46 | .0010030 | .9997299 |
| 3.07 | .0035836 | .9989297 | 3.27 | .0019010 | .9994623 | 3.47 | .0009689 | .9997398 |
| 3.08 | .0034751 | .9989650 | 3.28 | .0018397 | .9994810 | 3.48 | .0009358 | .9997493 |
| 3.09 | .0033695 | .9989992 | 3.29 | .0017803 | .9994991 | 3.49 | .0009037 | .9997585 |
| 3.10 | .0032668 | .9990324 | 3.30 | .0017226 | .9995166 | 3.50 | .0008727 | .9997674 |
| 3.11 | .0031669 | .9990646 | 3.31 | .0016666 | .9995335 | 3.51 | .0008426 | .9997759 |
| 3.12 | .0030698 | .9990957 | 3.32 | .0016122 | .9995499 | 3.52 | .0008135 | .9997842 |
| 3.13 | .0029754 | .9991260 | 3.33 | .0015595 | .9995658 | 3.53 | .0007883 | .9997922 |
| 3.14 | .0028835 | .9991553 | 3.34 | .0015084 | .9995811 | 3.54 | .0007581 | .9997999 |
| 3.15 | .0027943 | .9991836 | 3.35 | .0014587 | .9995959 | 5.55 | .0007317 | .9998074 |
| 3.16 | .0027075 | .9992112 | 3.36 | .0014106 | .9996103 | 3.56 | .0007061 | .9998146 |
| 3.17 | .0026231 | .9992378 | 3.37 | .0013639 | .9996242 | 3.57 | .0006814 | .9998215 |
| 3.18 | .0025412 | .9992636 | 3.38 | .0013187 | .9996376 | 3.58 | .0006575 | .9998282 |
| 3.19 | .0024615 | .9992886 | 3.39 | .0012748 | .9996505 | 3.59 | .0006343 | .9998347 |
| 3.20 | .0023841 | .9993129 | 3.40 | .0012322 | .9996631 | 3.60 | .0006119 | .9998409 |

* Biometrika Trustees এর অনুমতিক্রমে Biometrika Tables for Statisticians, Vol. 1 এর Table 1 থেকে সংক্ষেপিত আকারে মুদ্রিত।

## সারণী 2

### মৌল নর্মাল বিভাজন

( $\tau_\alpha$-এর কয়েকটি মান )

| $\alpha$ | 0.05 | 0.025 | 0.01 | 0.005 |
|---|---|---|---|---|
| $\tau_\alpha$ | 1.645 | 1.960 | 2.326 | 2.576 |

# নির্ঘণ্ট

# শুদ্ধিপত্র

| পৃষ্ঠা | লাইন | অশুদ্ধ অংশ (যা আছে) | শুদ্ধ অংশ (যা হবে) |
|---|---|---|---|
| 164 | 2 | $\overset{n}{\underset{i=1}{\cup}}$ | $\overset{n}{\underset{i=1}{\cup}} A_i$ |
| 167 | 7 | $P[A\cap B] + [A\cap B^*]$ | $P([A\cap B]+[A\cap B^*])$ |
| 170 | 3-এর পর | — | $\sum_{1}^{m+1} P(A_i) - \sum_{i<j}^{m+1} P(A_i\cap A_j) + \cdots + (-1)^m P(A_1\cap A_2\cap\cdots\cap A_{m+1})$ |
| 207 | 13, 21 | $\varepsilon$ | ε |
| 253 | 12 | $-\int \overset{v}{\cup}$ | $-\int_0^\infty$ |
| 291 | 14 | $\sum_{j=1}^{x}$ | $\sum_{j=1}^{s}$ |
| 295 | 14 | $f_{A\alpha O}$ | $f_{A\beta C}$ |
| ” | 16 | $(f_{AB\gamma}+f_{AB\gamma})$ | $(f_{AB\gamma}+f_{A\beta\gamma})$ |
| 296 | 4 | $f_{\alpha\beta}$ ও | $f_{\alpha B}$ ও |
| ” | 7 | $\theta$ | $\beta$ |
| ” | 8 | $f_{A\beta}+f_{\alpha\beta}=f_\alpha$ | $\cdots = f_\beta$ |
| ” | 16 | $\frac{f_{A\beta}}{f_B}$ | $\frac{f_{A\beta}}{f_\beta}$ |
| ” | 23 | $\frac{f_{AB}}{f_\beta}$ | $\frac{f_{A\beta}}{f_\beta}$ |
| ” | 24 | $(f_{AB}+f_{AB})$ | $(f_{AB}+f_{A\beta})$ |
| 300 | 26 | পুরোটাই | $f_{AB}(f_{AB}+f_{\alpha B}+f_{A\beta}+f_{\alpha\beta}) - (f_{AB}+f_{A\beta})(f_{AB}+f_{\alpha B})$ |